

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 07:34:31 : Search time 26 Seconds
(without alignments)
194.620 Million cell updates/sec

Title: US-09-733-685-2
Perfect score: 623
Sequence: 1 MNSLKKERVEDNCKSG.....QGLGLDNCKPEPDSVLSL 122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result, being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	80.5	12.9	968	PKD2_HUMAN	Q13563 homo sapien
2	77	12.4	966	PKD2_MOUSE	Q35245 mus musculus
3	76	12.2	1007	Y741_CHLMO	Q9P166 chlamydia m
4	72.5	11.6	448	YNQ6_YEAST	P53890 saccharomyc
5	72.5	11.6	493	GALT_STRSL	Q8V592 streptococc
6	72	11.6	444	CHLB_CHLPT	P37824 chlamydomon
7	71.5	11.5	543	THSB_TREAC	P48425 thermoplasm
8	71	11.4	217	GRE1_HUMAN	Q9HAV7 homo sapien
9	71	11.4	879	RA50_SULIO	Q96VR5 sulfolobus
10	70.5	11.3	301	TRT2_CHICK	P02642 gallus gall
11	70.5	11.3	775	RIR1_VZVD	P09248 varicella-z
12	70	11.2	320	IF34_HUMAN	Q75821 homo sapien
13	69	11.1	149	SP17_MOUSE	Q62252 mus musculu
14	69	11.1	843	CYP1_BRUMA	Q27450 brugia mala
15	69	11.1	2319	AKA6_HUMAN	Q13023 homo sapien
16	68.5	11.0	245	TRJA_THEMA	Q9X1R0 thermotoga
17	68.5	11.0	569	Y105_DROME	Q9X719 drosophila
18	68.5	11.0	837	RA53_METH	C28640 methanobact
19	68.5	11.0	1167	CAGA_HELPJ	Q9Z1T1 helicobacte
20	68	10.9	320	IF34_MOUSE	Q9Z1D1 mus musculu
21	68	10.9	499	UDPG_YEAST	P32861 saccharomyc
22	68	10.9	723	MY5B_MOUSE	P21271 mus musculu
23	68	10.9	1539	Y373_HUMAN	P15078 homo sapien
24	67.5	10.8	105	BL24_THEMA	P38513 thermotoga
25	67.5	10.8	240	SNF1_YEAST	P39929 saccharomyc
26	67.5	10.8	513	CATB_PSEAE	Q59635 pseudomonas
27	67.5	10.8	864	RA5C_SULSO	Q97WH0 sulfolobus
28	67	10.8	153	SCDC_DJBHA	Q42724 debaryomyce
29	67	10.8	562	CHS5_CANAL	Q74161 candida alb
30	67	10.8	596	Y41J_RHISN	P55493 rhizobium s
31	66.5	10.7	542	TULL_HUMAN	Q00294 homo sapien
32	66.5	10.7	569	YX1D_BACSU	P42295 bacillus su
33	66.5	10.7	568	YM18_YEAST	Q04511 saccharomyc

RESULT 1
PKD2_HUMAN
ID PKD2_HUMAN STANDARD: FRT: 968 AA.
AC Q13563: 060441: Q135754:
ST 15-JUL-1999 (Rel. 38, Created)
ST 15-JUL-1999 (Rel. 38, Last sequence update)
BT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polycystin 2 (Autosomal dominant polycystic kidney disease type II protein) (Polycystin) (R48321).
GN PKD2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RF SEQUENCE FROM N.A.
RA Mochizuki T., Wu G., Hayashi T., Xenophontos S.L., Veldhuisen B., Saris J.J., Reynolds D.M., Cai Y., Gabow P.A., Pierides A., Kimberling W.J., Breuning M.H., Deltas C.C., Peters D.J.M., Somlo S.; *PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein.*
RT Science 272:1339-1342(1996).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-97432829; PubMed-9286709;
RA Hayashi T., Mochizuki T., Reynolds D.M., Wu G., Cai Y., Somlo S.; *Characterization of the exon structure of the polycystic kidney disease 2 gene (PKD2).
RT Genomics 44:131-136(1997).
RL [3]
RP SEQUENCE OF 361-968 FROM N.A.
RC TISSUE=Breast;
RX Schneider M.C., Rodriguez A., Nomura H., Zhou J., Morton C.C., Reeders S.T., Wereniewicz S.; *A gene similar to PKD1 maps to chromosome 4q22: a candidate gene for PKD2.*
RT Genomics 38:1-4(1996).
RN [4]
RP VARIANT ADPKD GLY 414.
RX MEDLINE-97465600; PubMed-9326320;
RA Veldhuisen B., Saris J.J., de Haij S., Hayashi T., Reynolds D.M., Mochizuki T., Elies R., Fossdal R., Bogdanova N., van Dijk M.A., Coto E., Ravine D., Noerby S., Verellen-Dumoulin C., Breuning M.H., Somlo S., Peters D.J.M.; *A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2).
RT Am. J. Hum. Genet. 61:547-555(1997).
RN [5]
RP VARIANT ADPKD PRO-356 AND VARIANT PRO-28.
RX MEDLINE-99340490; PubMed-10411676;
RA Toira K., Virilbay M., Telleria D., Badenas C., Watson M., Harris P.C., Darnell A., San Millan J.L.; *Seven novel mutations of the PKD2 gene in families with autosomal dominant polycystic kidney disease.*
RT

ALIGNMENTS

34	66.5	10.7	681	1	NOGL_CAEEL	044411 caccorhabdi
35	66.5	10.7	866	1	MYSP_SCHJA	Q05670 schistosoma
36	65.5	10.5	399	1	HBAL_SCHPO	Q09146 schizosacch
37	65.5	10.5	405	1	PI52_MOUSE	P07172 mus musculu
38	65	10.4	237	1	COAT_TOBSV	P30598 tobacco str
39	65	10.4	527	1	TCPB_YEAST	P39076 saccharomyc
40	65	10.4	1040	1	EG15_CAEEL	Q10656 caenorhabdi
41	64.5	10.4	346	1	YE77_METJA	Q58872 methanococc
42	64.5	10.4	381	1	E28B_YEAST	P32502 saccharomyc
43	64.5	10.4	398	1	THIL_YEAST	P41338 saccharomyc
44	64.5	10.4	406	1	PI52_HUMAN	P08426 homo sapien
45	64.5	10.4	466	1	SYC_RHIME	Q92R20 rhizobium c

Kidney Int. 56:28-33(1999).

[6]

RP VARIANTS ADPKD ILE-479 DEL: 504-ARG--VAL-512 DEL AND -YR-684 DEL.

RX MEDLINE-20296613; PubMed-10835625;

RA Matlick T.J., He N., Wang K., Liang Y., Parfrey P., Hofferton D.,

RA St George-Hyslop P., Germino G.G., Pei Y.,

RT "Mutations of PKD1 in ADPKD2 cysts suggest a pathogenic effect of

RT trans-heterozygous mutations."

RL Nat. Genet. 25:143-144(2000).

CC -!- FUNCTION: PKD1 AND PKD2 MAY FUNCTION THROUGH A COMMON SIGNALING

CC PATHWAY THAT IS NECESSARY FOR NORMAL TUBULOGENESIS.

CC -!- SUBUNIT: INTERACTS WITH PKD1. PKD1 REQUIRES THE PRESENCE OF PKD2

CC FOR STABLE EXPRESSION.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN OVARY, FETAL AND ADULT

CC KIDNEY, TESTIS, AND SMALL INTESTINE. NOT DETECTED IN PERIPHERAL

CC LEUKOCYTES.

CC -!- DOMAIN: THE C-TERMINAL IS IMPLICATED IN THE INTERACTION WITH PKD1.

CC -!- DISEASE: DEFECTS IN PKD2 ARE THE CAUSE OF AUTOSOMAL DOMINANT

CC POLYCYSTIC KIDNEY DISEASE TYPE II (ADPKD TYPE II), WHICH REPRESENT

CC APPROXIMATELY 15% OF CASES OF AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY

CC DISEASE (ADPKD); A COMMON AUTOSOMAL DOMINANT GENETIC DISEASE

CC AFFECTING ABOUT 1 OUT 1000 INDIVIDUALS. IT IS CHARACTERIZED BY

CC PROGRESSIVE FORMATION AND ENLARGEMENT OF CYSTS IN BOTH KIDNEYS,

CC TYPICALLY LEADING TO END-STAGE RENAL DISEASE IN ADULT LIFE. CYSTS

CC ALSO OCCURS IN THE LIVER AND OTHER ORGANS. ALL MUTATIONS,

CC SCATTERED BETWEEN EXONS 1 AND 11, RESULT IN A TRUNCATED PKD2 THAT

CC LACKS BOTH THE CALCIUM-BINDING EF-HAND DOMAIN AND THE TWO

CC CYTOPLASMIC DOMAINS REQUIRED FOR THE INTERACTION OF PKD2 WITH PKD1

CC AND WITH ITSELF. PKD2 IS CLINICALLY Milder THAN PKD1, BUT IT HAS A

CC DELETERIOUS IMPACT ON OVERALL LIFE EXPECTANCY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to licen@isb-sib.ch).

CC

DR EMBL; U50928; AAC50520.1; .

DR EMBL; AF004873; AAC16004.1; JOINED.

DR EMBL; AF004859; AAC16004.1; JOINED.

DR EMBL; AF004860; AAC16004.1; JOINED.

DR EMBL; AF004861; AAC16004.1; JOINED.

DR EMBL; AF004862; AAC16004.1; JOINED.

DR EMBL; AF004863; AAC16004.1; JOINED.

DR EMBL; AF004864; AAC16004.1; JOINED.

DR EMBL; AF004865; AAC16004.1; JOINED.

DR EMBL; AF004866; AAC16004.1; JOINED.

DR EMBL; AF004867; AAC16004.1; JOINED.

DR EMBL; AF004868; AAC16004.1; JOINED.

DR EMBL; AF004869; AAC16004.1; JOINED.

DR EMBL; AF004870; AAC16004.1; JOINED.

DR EMBL; AF004871; AAC16004.1; JOINED.

DR EMBL; AF004872; AAC16004.1; JOINED.

DR EMBL; U56813; AAC50333.1; .

DR Gene; HGNC:9009; PKD2.

DR MIM; 173910; .

DR InterPro; IPR001582; Ca/Na_pore.

DR InterPro; IPR002111; Cal_channel_TrpL.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000636; M-channel_nig.

DR InterPro; IPR003915; PKD_2.

DR Pfam; PF00520; ion_trans_1.

DR PRINTS; PR01433; POLYCYSTIN2.

DR PROSITE; PS00018; EF_HAND; FALSE_NEG.

KW Transmembrane; Glycoprotein; Coiled coil; Calcium-binding;

KW Disease mutation; Polymorphism.

FT DOMAIN 1 223 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 224 244 POTENTIAL.

FT DOMAIN 245 468 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 469 489 POTENTIAL.

FT

505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
MEDLINE=98384541; PubMed=9716661;
Pennekamp P., Boydanova N., Wilda M., Markoff A., Hameister H.,
Forst J., Dworniczak B.,
"Characterization of the murine polycystic kidney disease (Pkd2)
gene".
Mamm. Genome 9:749-752(1998).
-!- FUNCTION: PKD1 AND PKD2 MAY FUNCTION THROUGH A COMMON SIGNALING
PATHWAY THAT IS NECESSARY FOR NORMAL TUBULOGENESIS (BY
SIMILARITY).
-!- SUBUNIT: INTERACTS WITH PKD1. PKD1 REQUIRES THE PRESENCE OF PKD2
FOR STABLE EXPRESSION (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
-!- TISSUE SPECIFICITY: EXPRESSED IN MESENCHYMALLY DERIVED STRUCTURES
IN THE DEVELOPING EMBRYO AT DAY 12.5. IN ADULT, MOSTLY EXPRESSED
IN KIDNEY.
-!- DOMAIN: THE C-TERMINAL IS IMPLICATED IN THE INTERACTION WITH PKD1
(BY SIMILARITY).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF014010; AAC53388.1; -
EMBL: Y13278; CAA73727.1; -
EMBL: Y14105; CAA74551.1; -
EMBL: Y14106; CAA74551.1; JOINED.
EMBL: Y14107; CAA74551.1; JOINED.
EMBL: Y14108; CAA74551.1; JOINED.
EMBL: Y14109; CAA74551.1; JOINED.
EMBL: Y14110; CAA74551.1; JOINED.
EMBL: Y14111; CAA74551.1; JOINED.
EMBL: Y14112; CAA74551.1; JOINED.
EMBL: Y14113; CAA74551.1; JOINED.
EMBL: Y14114; CAA74551.1; JOINED.
EMBL: Y14115; CAA74551.1; JOINED.
EMBL: Y14116; CAA74551.1; JOINED.
EMBL: Y14117; CAA74551.1; JOINED.
EMBL: Y14118; CAA74551.1; JOINED.
EMBL: Y14119; CAA74551.1; JOINED.
EMBL: Y14120; CAA74552.1; -
HSP: P02632; ICBI.
MGD: MGI:1099818; Pkd2.
InterPro: IPR001682; Ca/Na_pore.
InterPro: IPR002111; Cat_channel_TrpL.
InterPro: IPR002048; EF-hand.
InterPro: IPR000636; M+channel_nig.
InterPro: IPR003915; PKD_2.
Frame: P00520; ion_trans...
PRINTS: PR01433; POLYCYSTIN2.
PROSITE: PSC0018; EF_HAND; FALSE_NEG.
Transmembrane: Glycoprotein; Coiled coil; Calcium-binding.
CYTOPLASMIC (POTENTIAL).
DOMAIN 1 221
TRANSMEM 222 242
DOMAIN 243 466
TRANSMEM 467 487
DOMAIN 488 503
TRANSMEM 504 524
DOMAIN 525 548
TRANSMEM 549 569
DOMAIN 570 596
TRANSMEM 597 617
DOMAIN 618 656
TRANSMEM 657 677
DOMAIN 678 966
CYTOPLASMIC (POTENTIAL).
POLY-ASP.
DOMAIN 95 99
DOMAIN 151 154
CA_BIND 761 772
DOMAIN 761 794
COILED COIL (POTENTIAL).

FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 365 365 I -> M (IN REF. 2).
FT CONFLICT 370 370 R -> K (IN REF. 2).
FT CONFLICT 560 560 A -> S (IN REF. 2).
FT CONFLICT 688 688 DL -> SV (IN REF. 2; CAA74551).
FT CONFLICT 746 746 K -> E (IN REF. 2).
FT CONFLICT 942 942 S -> P (IN REF. 2).
FT CONFLICT 957 957 S -> G (IN REF. 2).
SQ SEQUENCE 966 AA: 108995 MW: 888888889935F CRC64;
Query Match 12.4%; Score 77; DB 1; Length 966;
Best Local Similarity 29.1%; Pred. No. 12;
Matches 30; Conservative 17; Mismatches 44; Indels 12; Gaps 3;
QY 8 FERVEESNKGSDNCRKPTSTEVVRIVTEEEVDFKILRRV----HVATRIYAKVNGVA 63
DY 811 FEUDDSDGSHSRHRSISGV-----SYEEFQVLRVVRVDMHSHSGSIYSKIDAVIV 96;
QY 64 EDELPRK-KRKRSONGLKNSLDCNGVRDGEFEINRVGLQL 105
DY 664 KLSIMPRALKPREVLGRLLDGVAEADARLGRDSEIHREQMERL 906

RESULT 3
Y741_CHLMU
ID Y741_CHLMU STANDARD; PRI: 1007 AA.
AC Q9PJT6;
DI 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TC0741 precursor.
GN TC0741.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
Linher K., Weidman J., Kouri H., Craven B., Bowman C., Dodson R.,
Swinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
Nucleic Acids Res. 28:1397-1406(2000).
RT -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0572/CT456/TC0741
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: AE002342; AAF39550.1; -
DR TIGR: TC0741; -
DR Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 51 51 POTENTIAL.
FT CHAIN 52 1007 HYPOTHETICAL PROTEIN TC0741.
SQ SEQUENCE 1007 AA: 104906 MW: 8428000871518 CRC64;
Query Match 12.2%; Score 76; DB 1; Length 1007;
Best Local Similarity 32.2%; Pred. No. 15;
Matches 29; Conservative 11; Mismatches 32; Indels 18; Gaps 5;

CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
 CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
 CC of three subunits: chl_a, chl_b and chl_c. Could form a
 CC heterotetramer of two chl_b and two chl_c subunits (by similarity).
 CC -!- SIMILARITY: BELONGS TO THE CHL_B / BCH₂ / BCH₂ FAMILY.
 DR InterPro: IPR000510; Oxred_nitrogsnel.
 DR Pfam: PF00148; oxidored_nitro_1.
 DR Oxidoreductase; Photosynthesis; Chlorophyll biosynthesis; Chloroplast.
 DR NON_TER 444 444
 DR SEQUENCE 444 AA: 50058 MW: 650652CAAE4857F CRC64;
 Query Match 11.6%; Score 72; DB 1; Length 444;
 Best Local Similarity 25.0%; Pred. No. 14;
 Matches 32; Conservative 17; Mismatches 44; Indels 30; Gaps 6;
 2Y 11 VEDNGKSDGNRGKSTEVRTVTEEEVDEFFKILRRVHVHVAITRVAKVNGVAGE--- 56
 b 113 TSESNHGSGIDKATSDVILADVNNHYRNE-----LQADRTLEGIVRFYLEKEKLN 165
 Y 57 ---LPSKRRKRSQN-----LGLNSLDCNGVRDGEFDEINRVGLQGLGLEN-CKPEPD 226
 b 166 INTPTK-KKPSANTIGFTISFHNQDCR-----ELKRL-ENNIG-EVNEIIPGG 216
 Y 117 SVS 119
 b 217 SVT 219
 RESULT 7
 HSB_THEAC
 D THSB THEAC STANDARD; PRT: 543 AA.
 C P48425;
 T 01-FEB-1996 (Rel. 33, Created)
 T 01-FEB-1996 (Rel. 33, Last sequence update)
 T 16-OCT-2001 (Rel. 40, Last annotation update)
 E Thermosome beta subunit (thermosome subunit 2) (Chaperonin beta
 E subunit).
 N THSB OR TA1276.
 S Thermoplasma acidophilum.
 S Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatiales;
 C Thermoplasmataceae; Thermoplasma.
 X NCBI_TaxID=2303;
 N [1]
 P SEQUENCE FROM N.A., AND SEQUENCE OF 188-195.
 C STRAIN-ATCC 25905;
 X MEDLINE=95314774; PubMed=7794526;
 A Waldmann T., Lupas A.N., Kellermann J., Peters J., Baumeister W.;
 T "Primary structure of the thermosome from Thermoplasma acidophilum.";
 T Biol. Chem. Hoppe-Seyler 376:119-126(1995).
 P [2]
 P SEQUENCE FROM N.A.
 C STRAIN-DSM 1728;
 X MEDLINE=20479972; PubMed=11029001;
 A Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 X Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 T "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 T acidophilum.";
 T Nature 407:508-513(2000).
 L [3]
 N SEQUENCE OF 80-113 AND 421-445.
 P STRAIN-ATCC 25905;
 X MEDLINE=95172073; PubMed=7867646;
 A Waldmann T., Nimesgern E., Nitsch M., Peters J., Pfeifer G.,
 X Mueller S., Kellermann J., Engel A., Hartl F.-O., Baumeister W.;
 T "The thermosome of Thermoplasma acidophilum and its relationship to
 T the eukaryotic chaperonin TRiC.";
 L Eur. J. Biochem. 227:848-856(1995).
 N [4]
 P X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
 X MEDLINE=98206474; PubMed=9546398;
 A Ditzel L., Lowe J., Stock D., Stetter K.O., Huber H., Huber R.,
 A Steinbacher S.;
 T "Crystal structure of the thermosome, the archaeal chaperonin: and

RT homolog of OCT.";
 RL Cell 93:125-138(1998).
 CC -!- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN
 CC V-TRO. AND HAS A WEAK ATPASE ACTIVITY.
 CC -!- SUBUNIT: FORMS A HETERO-OLIGOMERIC COMPLEX OF TWO STACKED EIGHT-
 CC MEMBERED RINGS.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL: Z46650; CAA86611.1;
 DR EMBL: A2445067; CAC22400.1;
 DR PDB: 1A6E; 23-MAR-99.
 DR PDB: 1A6E; 23-MAR-99.
 DR InterPro: IPR002194; Chaperonin_TCP-1.
 DR InterPro: IPR002423; Cpn60/TCP-1.
 DR Pfam: PF00118; cpn60_TCP1.1.
 DR PROSITE: PS00750; TCP1_1.1.
 DR PROSITE: PS00751; TCP1_2.1.
 DR PROSITE: PS00955; TCP1_3.1.
 KW Chaperone; ATP-binding; 3D-structure; Complete proteome.
 KW SEQUENCE 543 AA: 58479 MW: F9DAEE63EB84E6 CRC64;
 Query Match 11.5%; Score 71.5; DB 1; Length 543;
 Best Local Similarity 26.9%; Pred. No. 20;
 Matches 32; Conservative 15; Mismatches 59; Indels 13; Gaps 5;
 QY 2 NNSLKEERVEEDNGKSD-----GNRGKSTEV-VRVTVEEVDDEFFK-ILRRVHVHVAIR 53
 b 339 SSDLGTAERVEQVKVEDYMTFTVTCCKNPKAVSILVRGETEIVHVDMSERTDLSLHVAS 398
 QY 54 TVA-----KVGSGVAGELPSKRRKRSQNLGNSLDCNGVRDGEFDEINRVGLQGLGLD 108
 b 399 ALPDGAYAGGGATAAEIAFLRSYAKIGGQQQLAIEKFADA-IEEIPRALAENAGLD 456
 RESULT 8
 GREI_HUMAN
 ID GREI_HUMAN STANDARD; PRT: 217 AA.
 AC Q9HAV7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GrpE protein homolog 1, mitochondrial precursor (Mt-GrpE1) (HMG).
 GN GREP1...
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 2-117 FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Brain;
 RX MEDLINE=21211408; PubMed=11311562;
 RA Choglay A.A., Chapple J.P., Blatch G.L., Cheetham M.E.;
 RT "Identification and characterization of a human mitochondrial
 RL homologue of the bacterial co-chaperone GrpE.";
 RN Gene 267:125-134(2001).
 RP SEQUENCE FROM N.A.
 RA Li W.B., Graber C., Jessee J., Polayes J.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

```

CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an Atp-dependent control of mreII by unwinding
CC and/or repositioning DNA ends into the mreII active site (By
CC similarity).
CC
CC -|- SUBUNIT: Forms a complex with mreII (By similarity).
CC
CC -|- SIMILARITY: BELONGS TO THE SMC FAMILY, RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF000988; BAB672.2.1: ALT_INIT.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR003405; SMC_C.
CC InterPro: IPR003395; SMC_N.
CC Pfam: PF00005; ABC_tran.1.
CC Pfam: PF02483; SMC_C.1.
CC Pfam: PF02463; SMC_N.1.
CC
CC Problem: P6000006; ABC_transporter.1.
CC DNA repair; Hydrolase; Atp-binding; Coiled coil;
CC NP_BIND 30 37 Atp (By similarity).
CC Complete proteome.
CC
CC KW COILED COIL (POTENTIAL).
CC FT DOMAIN 72 731
CC SEQUENCE 879 AA: 103487 MW: 89103858914F20D7 CRC64:
CC
CC Query Match 11.4%; Score 71; DB 1; Length 879;
CC Best Local Similarity 20.8%; Pred. No. 38;
CC Matches 25; Conservative 26; Mismatches 29; Indels 40; Gaps 5;
CC
CC QY 1 MNNSLKKEP-----RVEEDN-----GK--SDGNRGKPSIEVV 30
CC
CC DQ 401 LNNKNGERKIASQLCEKGRIEELKLLGNLVGNVCPVCGRELSDDHKRKGIONELI 460
CC
CC QY 31 RTVTE-EEYDEFFKILRVHVRVTRVAKVNGVGAEGELPSKKRKRSONGLRNSLDCNV 89
CC
CC DQ 461 EKLKEDLNNKFKL-----EINKINGLISELNQIINKSKKEDIARNLADYNKL 511
CC
CC
CC RESULT 10
CC TRF2_CHICK
CC ID TRF2_CHICK STANDARD: PR7: 301 AA.
CC AC P02642;
CC DT 21-JUL-1986 (Rel. 51, Created)
CC DT 01-JUN-1988 (Rel. 58, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Troponin T, cardiac muscle isoforms (TnTC).
CC OS Gallus gallus (Chicken).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
CC OC Gallus.
CC
CC QN NCBI_TaxID=9031;
CC QX [1];
CC RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
CC RX MEDLINE=85289327; PubMed=2953302;
CC RA Cooper T.A., Ordahl C.P.;
CC RT "A single cardiac troponin T gene generates embryonic and adult
CC isoforms via developmentally regulated alternate splicing."
CC RL J. Biol. Chem. 260:11140-11146(1985).
CC RN [2];
CC RP SEQUENCE OF 67-301 FROM N.A.
CC RX MEDLINE=85065747; PubMed=6095446;
CC RA Cooper T.A., Ordahl C.P.;
CC RT "A single troponin T gene regulated by different programs in cardiac
CC and skeletal muscle development."
CC RL Science 226:979-982(1984)
CC
CC -|- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC TROPONIN, THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS
CC CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.
CC
CC -|- ALTERNATIVE PRODUCTS: 2 isoforms; 1/Embryonic form (shown here)

```

```

CC      and 2/adult form; are produced by alternative splicing.
CC      -!- SIMILARITY: BELONGS TO THE TROPONIN T FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M10013; AAA49099.1;
CC      EMBL: K02263; AAA49098.1;
CC      PIR: A03086; TPCHPC.
CC      PIR: A25373; A25373.
CC      InterPro: IPR001978; Troponin.
CC      Pfam: PF00992; Troponin; i.
CC      Muscle protein; Alternative splicing; Multigene family;
CC      Phosphorylation.
CC      INIT_MET 0 0 BY SIMILARITY.
CC      MOD_RES 1 1 PHOSPHORYLATION (BY CK2)
CC      (BY SIMILARITY).
CC      VARSPLIC 22 31 MISSING (IN ISOFORM 2).
CC      SEQUENCE 301 AA; 35854 MW; F85CEHIA47F07DD94 CRC64;
CC      -----
CC      Query Match: 11.3%; Score 70.5; DB 1; Length 301;
CC      Best Local Similarity 28.7%; Pred. No. 13;
CC      Matches 33; Conservative 13; Mismatches 36; Indels 33; Gaps 6;
CC      -----
2y 6 KSERVEEDNGKSGRGRKSTVETVEEVDKFKILRRVHVATRVKAVNGS--- 51
db 22 EEEVLEEDGGQED-QVDEEESTETTTAEQDE-----TKANGEGEGSR 67
y 62 ---VAEGELSKKSKKRSQNLGRLNSLDCNGVRDGE---FDEINRVGLOGGLDNL 110
b 68 EQEPGEGESKPKPKPMPNLPVPPK-----LPDGERLDFDDTHRKWKE---KDLN 113
-----
RESULT 11
IRI_VZVD
D RIRL_VZVD STANDARD: PRT: 775 AA.
C P09248;
T 01-MAR-1989 (Rel. 10, Last: sequence update)
T 01-MAR-1989 (Rel. 10, Last: sequence update)
T 01-JUN-1994 (Rel. 29, Last: annotation update)
E Ribonucleoside-diphosphate reductase large chain (EC 1.1.7.4.1)
E (Ribonucleotide reductase).
N 19.
S Varicella-zoster virus (strain Dumas) (VZV).
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Varicellovirus.
X NCBI_TaxID=10338;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=86306657; PubMed=2018124;
A Davison A.J., Scott J.E.;
T "The complete DNA sequence of varicella-zoster virus.";
J. Gen. Virol. 67:1759-1816(1986).
C -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
C -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
C thioredoxin + H(2)O -> ribonucleoside diphosphate + reduced
C thioredoxin.
C -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
C -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
C -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE
C LARGE CHAIN FAMILY.
C -----
C This SWISS-PROT entry is copyrighted. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announcement/
C or send an email to license@isb-sib.ch).
C -----

```

```

CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X04370; CAA27902.1;
CC      PIR: A27343; WMBE19.
CC      InterPro: IPR000788; Ribonucleo_red.
CC      Pfam: PF00317; ribonuc_red_lg; 1.
CC      Pfam: PF02867; ribonuc_red_lgC; 1.
CC      PRINTS: PR01183; RIBORPTASEM1.
CC      PROSITE: PS00089; RIBORED_LARGE; 1.
CC      Oxidoreductase, DNA replication; Early protein.
CC      SEQUENCE 775 AA; 86827 MW; 181695A65E891B9D CRC64;
CC      -----
CC      Query Match: 11.3%; Score 70.5; DB 1; Length 775;
CC      Best Local Similarity 31.5%; Pred. No. 37;
CC      Matches 23; Conservative 7; Mismatches 14; Indels 29; Gaps 4;
CC      -----
2y 51 AFTVAKVNGVGAEGELPSKKKRSQNLGRLNSLDCNGVRDGEFDEINR---VGLQG--- 104
db 471 ARTVIFLNGVLAAGNPPCKK-----SKGVKN-----NRSLGIGIGG:HT 513
y 105 ---LGLDLNCKP 113
b 512 TCLRIGFDLTSP 524
-----
RESULT 12
IF34_HUMAN
ID IF34_HUMAN STANDARD: PRT: 320 AA.
AC Q75921; O14801;
CT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last: sequence update)
DT 15-JUN-2002 (Rel. 41, Last: annotation update)
DE Eukaryotic translation initiation factor 3 subunit 4 (eIF3 delta)
DE (EIF3 p44) (eIF-3 RNA-binding subunit) (EIF3 p42).
GN EIF3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99041954; PubMed=9822859;
RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human
RT translation initiation factor eIF3.";
RL J. Biol. Chem. 273:31901-31908(1998).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=99141230; PubMed=9973622;
RA Bandhopadhyay A., Maitra U.;
RT "Cloning and characterization of the p42 subunit of mammalian
RT translation initiation factor 3 (eIF3): demonstration that eIF3
RT interacts with eIF5 in mammalian cells.";
RL Nucleic Acids Res. 27:1331-1337(1999).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA. THIS SUBUNIT BINDS TO THE 18S RRNA.
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U96074; AAC78728.1;
CC EMBL: AF020833; AAB71866.1;
CC HSP: P19339; 2SXL.
CC Genew: HGNC:3274; EIF3A4.
CC MIM: 603913;
CC InterPro: IPR000504; RNA_rec_mot.
CC -----

```

```
DR Pfam: PF00076; trm: 1.
DR SMART: SM00360; RRM: 1.
DR PROSITE: PS00102; RRM: 1.
DR PROSITE: PS00300; RRM_RNF_1: 1.
DR Initiation factor: Protein biosynthesis: RNA-binding.
DR DOMAIN 239 317 RNA-BINDING (SRM).
DR FT CONFLICT 293 293 K -> A (IN REF. 2).
DR SEQUENCE 320 AA: 35696 MW: 70722626CECFB65FB6 CRC64;

Query Match 11.1%; Score 70; DB 1; Length 320;
Best Local Similarity 26.1%; Pred. No. 15;
Matches 30; Conservative 17; Mismatches 26; Indels 42; Gaps 7;

QY 25 PSTEV-----VRTVTEEEVDE---PFKILRRVIVATRTVAKVNGGVAEGFLPSKKKKRSCN 77
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 55 PKKEVINQKVTVEIKDECKKFKIVATRIETRKASK---AVA-----RKNKKK 104
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 78 LQLRNSLDCNGVRDEP-----DEINRVGLQGLDGNKPRPDSVS 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 105 PG-----NSEFDPGPNVATTVSDVVSMTFTISKF-DLNCQEEEDPMN 147
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
SP17_MOUSE
ID Q62252; STANDARD; PRT: 149 AA.
AC Q62252;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sperm surface protein Sp17 (Sperm autoantigenic protein 17).
DN SPAL7 OR SP17.
JC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96039129; PubMed=7578682;
RA Kong M., Richardson R.T., Wigdore E.E., O'Rand M.G.;
RA "Sequence and localization of the mouse sperm autoantigenic protein,
RA Sp17.";
RA Biol. Reprod. 53:579-590(1995).
RL FUNCTION: SPERM SURFACE ZONA PELLUCIDA BINDING PROTEIN. HELPS TO
RL BIND SPERMATOZOEA TO THE ZONA PELLUCIDA WITH HIGH AFFINITY. MIGHT
RL FUNCTION IN BINDING ZONA PELLUCIDA AND CARBOHYDRATES (BY
RL SIMILARITY).
RL SUBUNIT: HOMODIMER (BY SIMILARITY).
RL TISSUE SPECIFICITY: TESTIS- AND SPERM-SPECIFIC.
RL SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: 245299; CA886455.1;
CC HSSP: P12367; IR2A.
CC MGD: MGI:133378; Spal7.
CC InterPro: IPR000048; IQ region.
CC InterPro: IPR003117; RIIa.
CC Pfam: PF00612; IQ: 1.
CC Pfam: PF02197; RIIa; 1.
CC SMART: SM00615; IQ: 1.
CC SMART: SM00394; RIIa; 1.
CC PROSITE: PS50096; IQ: 1.
CC Membrane. 112 141 IQ.
CC DOMAIN
CC SEQUENCE 149 AA: 17296 MW: C7E05D11D6AFDC CRC64;

Query Match 11.1%; Score 69; DB 1; Length 149;
Best Local Similarity 27.2%; Pred. No. 8.1;
Matches 25; Conservative 11; Mismatches 34; Indels 22; Gaps 3;

QY 2 NNSLKKKERY--EEONGKSDGNRGKPSSTEVTVTEEEVDEFFKILR-----RVHVATR 53
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 71 NHAFKEQGEVKECEQELAKSSGKRETFVTFPEESTTEERPOEEAAAKIKQSLFRGHVARE 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 54 TVAKVNGGVAEGELPSKKKKRSQNLGRNSLD 85
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 EV-----KKMKSDKNENLKEAD 148
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
CYP1_BROMA
ID CYP1_BROMA STANDARD; PRT: 843 AA.
AC Q27450;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptidyl:prolyl isomerase CYP-1 (EC 5.2.1.8) (Peptidylprolyl cis-trans
DE isomerase) (Cyclophilin) (PPIase).
DN CYP.
OC Brugia malayi.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Brugia.
OX NCBI_TaxID=6279;
RN [-],
RP SEQUENCE FROM N.A.
RA MEDLINE=96009642; PubMed=7547885;
RA Page A.P., Landry D., Wilson G.G., Carlow C.K.S.;
RA "Molecular characterization of a cyclosporin A-insensitive
RA cyclophilin from the parasitic nematode Brugia malayi.";
RA Biochemistry 34:11545-11550(1995).
RN [2],
RP X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS) OF 1-177.
RA MEDLINE=98218582; PubMed=9559680;
RA Taylor P., Page A.P., Kontopidis G., Husi H., Walkinshaw M.D.;
RA "The X-ray structure of a divergent cyclophilin from the nematode
RA parasite Brugia malayi.";
RA FEBS Lett. 425:361-366(1998).
RN [3],
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 1-177.
RA S-RAIN-ATCC 75593;
RA MEDLINE=98318040; PubMed=9655334;
RA Mikol V., Ma D., Carlow C.K.S.;
RA "Crystal structure of the cyclophilin-like domain from the parasitic
RA nematode Brugia malayi.";
RA Protein Sci. 7:1310-1316(1998).
RN [4],
RP X-RAY CRYSTALLOGRAPHY (2.47 ANGSTROMS) OF 1-177.
RA MEDLINE=20108543; PubMed=10642184;
RA Ellis P.J., Carlow C.K.S., Ma D., Kuhn P.;
RA "Crystal structure of the complex of Brugia malayi cyclophilin and
RA cyclosporin A.";
RA Biochemistry 39:592-598(2000).
RL FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
RL the cis-trans isomerization of proline imidic peptide bonds in
RL oligopeptides.
RL CATALYTIC ACTIVITY: Peptidylproline (omega-180) - peptidylproline
RL (omega=0).
RL ENZYME REGULATION: RELATIVELY INSENSITIVE TO INHIBITION BY CSA.
RL SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```



```

DR EMBL: L37292; AAC37249.1;
DR FDB: 1A58; 27-MAY-98.
DR PDB: 1A33; 29-JUL-98.
DR PDB: 1C5F; 02-DEC-99.
DR InterPro: IPR002130; CSA_PPIase.
DR Pfam: PF00160; PRO_Isomerase; 1.
DR PRINTS: PR00153; CSAPPISKRASE.
DR PROSITE: PS00170; CSA_PPIase_1; 1.
DR PROSITE: PS00072; CSA_PPIase_2; 1.
KW Isomerase; Rotamase; 3D-structure.
FT DOMAIN 10 175 PPIASE, CYCLOPHILIN-TYPE.
FT DOMAIN 700 709 POLY-ARG.
FT DOMAIN 713 716 POLY-ARG.
FT DOMAIN 800 815 POLY-SER.
FT DOMAIN 828 837 POLY-SER.
FT SEQUENCE 843 AA: 97817 MW: 30343C90A32EDBDC CRC64;

Query Match 11.1%; Score 69; DB 1; Length: 843;
Best Local Similarity 26.0%; Pred. No. 57;
Matches 26; Conservative 13; Mismatches 51; Indels 10; Gaps 2;

4Y 1 MNLSLKEERVEEDNCKSD--GNRQKPFSTEYVTVTEEEVDEFFKILRRVHVATRTVAKV 58
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4B 373 INTSEKIDKEERYKSKSKENRHSRSHHTTSPFHVTRHFVKEKNR-----HKV 424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

4Y 59 NGVAGGELPSKKRKSQNLGNLSLDCNVRDGEFDEIN 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4B 425 DEVGNSDMKQTRDRGRADEKEKVEVNGEKAAMDELN 464
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
AKAP6_HUMAN STANDARD: PRT: 2319 AA.
AC Q13023; Q15028;
T 16-OCT-2001 (Rel. 40, Created)
T 16-OCT-2001 (Rel. 40, Last sequence update)
T 15-JUN-2002 (Rel. 41, Last annotation update)
E A-kinase anchor protein 6 (Protein kinase A anchoring protein; 6)
E (PRK46) (A-kinase anchor protein 100 kDa) (AKAP 100) (MAKAP).
N AKAP6 OR AKAP100 OR KIAA0311.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
N [1]
N SEQUENCE FROM N.A.
X MEDLINE=99343692; PubMed=10413680;
A Kapiloff M.S., Shillace R.V., Westphal A.M., Scott J.D.:
T "MAKAP: an A-kinase anchoring protein targeted to the nuclear membrane
T of differentiated myocytes."
L J. Cell Sci. 112:2725-2736(1999).
N [2]
N SEQUENCE OF 934-2319 FROM N.A.
C TISSUE=Brain;
X MEDLINE=97349984; PubMed=9205841;
A Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
A Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.:
T "Prediction of the coding sequences of unidentified human genes. VII.
T The complete sequences of 100 new cDNA clones from brain which can
T code for large proteins in vitro."
L DNA Res. 4:141-150(1997).
N [3]
N SEQUENCE OF 1666-2319 FROM N.A.
C TISSUE=Hippocampus;
X MEDLINE=95238446; PubMed=7721854;
A McCartney S., Little B.M., Langeberg L.K., Scott J.D.:
T "Cloning and characterization of A-kinase anchor protein: 100
T (AKAP100). A protein that targets A-kinase to the sarcolemmal
T reticulum."
J. Biol. Chem. 270:9327-9333(1995).
-!- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE A
AND ANCHORS/TARGETS THEM TO THE NUCLEAR MEMBRANE OR SARCOPLASMIC
RETICULUM. MAY ACT AS AN ADAPTER FOR ASSEMBLING MULTIPROTEIN

```

```

CC COMPLEXES.
CC -!- SUBUNIT: INTERACTS WITH RII SUBUNIT OF PKA, PHOSPHATASE 2B
CC (CALCINEURIN) AND AKAP79.
CC -!- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM AND NUCLEAR MEMBRANE
CC IN HEART MUSCLE. PARTICIPATION OF MULTIPLE TARGETING SIGNALS ALLOW
CC CORRECT INTRACELLULAR TARGETING. THESE MAY BE REPEATED MOTIFS RICH
CC IN BASIC AND HYDROPHOBIC AMINO ACIDS, PALMITOYLATED/MYRISTOYLATED
CC MOTIFS OR ALTERNATIVELY SPLICE TARGETING SEQUENCES.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN CARDIAC AND SKELETAL
CC MUSCLE, FOLLOWED BY BRAIN.
CC -!- DOMAIN: RII-ALPHA BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC
CC HELIX, COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC -!- SIMILARITY: CONTAINS 2 SPECTRIN REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annuncce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U17195; AA032354.2;
CC EMBL: AH002309; BAA20770.1;
CC Genew: HGNC:376; AKAP6.
CC MIM: 604691;
CC InterPro: IPR002017; Spectrin.
CC Pfam: PF00435; Spectrin; 2.
CC SMART: SMC0150; SPEC; 2.
CC Repeat: 762 848 SPECTRIN 1.
CC REPEAT 1036 1150 SPECTRIN 2.
CC DOMAIN 1560 1701 SER-RICH.
CC CHAIN 2063 2076 PKA-RII SUBUNIT BINDING DOMAIN.
CC CONFLICT 974 974 C -> W (IN REF. 2).
CC CONFLICT 1492 1492 V -> A (IN REF. 2).
CC SEQUENCE 2319 AA: 256663 MW: 303037AE40FCFEE CRC64;

Query Match 11.1%; Score 69; DB 1; Length: 2319;
Best Local Similarity 25.8%; Pred. No. 1.8e+02;
Matches 32; Conservative 20; Mismatches 52; Indels 20; Gaps 7;

QY 1 MNLSLKEERVEEDNCKSDGNKPKFSTEYVTVTEEEVDE-FFKILRRVHVATRTVAKVN 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1849 LNSVK---RVSENG-----NGKNSSHHELGTREKNTKTKFYKVKOPYVADMENGNI 1900
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 GGVAEGE--LPSKKRKSQNLGNLS-----LDCNVRDGEFDEINRVGLGLDLN 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1901 -G:TERQKQKPNVTSKVSENLAGSHGKEISESEHCKCKALMD-SLDDSNATAGKEFYSDVR 1958
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 111 CKPE 114
      : :
DB 1959 HLPK 1962
      : :

Search completed: December 4, 2002, 09:23:17
Job time : 34 secs

```

GenCore version 5.1.3
 Copyright (C) 1993 - 2002 CompuGen, Ltd.
 OM protein - protein search, using sw model
 Run on: December 4, 2002, 08:27:54 : Search time 82 seconds
 (without alignments)
 306.558 Million cell updates/sec

Title: US-09-733-685-2
 Perfect score: 623
 Sequence: 1 MNKSLKKERVEEDNGKSDG.....QQLGELNCKPEPDSVLSL 122

Scoring table: BLOSUM62
 Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues
 Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database: SP-REMBL2.1:
 1: sp.archaea:
 2: sp.bacteria:
 3: sp.fungi:
 4: sp.human:
 5: sp.invertebrate:
 6: sp.mammal:
 7: sp.mhc:
 8: sp.orcanelle:
 9: sp.phage:
 10: sp.plant:
 11: sp.protozoa:
 12: sp.virus:
 13: sp.vertibrate:
 14: sp.unclassified:
 15: sp.virus:
 16: sp.bacteriap:
 17: sp.archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	623	100.0	122	10 Q9LJA3	Q9LJA3 Arabidopsis
2	96.5	15.5	85	10 Q82448	Q82448 Nicotiana t
3	80.5	12.9	265	16 Q92888	Q92888 chlamydia p
4	80.5	12.9	2254	10 Q9LNC2	Q9LNC2 Arabidopsis
5	80	12.8	753	2 Q9X553	Q9X553 neisseria m
6	79.5	12.8	2001	10 Q9M659	Q9M659 Arabidopsis
7	79.5	12.8	2001	10 Q9M658	Q9M658 Arabidopsis
8	78.5	12.6	265	16 Q9K2A4	Q9K2A4 chlamydia p
9	78	12.5	252	10 Q9PJF8	Q9PJF8 Arabidopsis
10	77.5	12.4	302	9 Q9TIE8	Q9TIE8 actinobacill
11	76.5	12.3	386	10 Q9FLD3	Q9FLD3 Arabidopsis
12	75.5	12.1	325	12 Q69272	Q69272 leporid her
13	75	12.0	354	2 Q9R3S1	Q9R3S1 actinobacill
14	75	12.0	365	2 Q9RAV2	Q9RAV2 actinobacill
15	75	12.0	375	2 Q9RCG2	Q9RCG2 actinobacill
16	75	12.0	404	5 Q95YR8	Q95YR8 leishmania

ID	Q9LJA3	PRELIMINARY:	PRT:	122 AA.
AC	Q9LJA3			
DT	01-OCT-2000 (TRENBLrel. 15, Created)			
DT	01-OCT-2000 (TRENBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TRENBLrel. 15, Last annotation update)			
DE	Genomic DNA, chromosome 3, P1 clone: MPE11 (NIMIN-2 protein).			
GN	NIMIN-2.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
CX	NCBI TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=COLUMBIA;			
RA	Sato S., Nakamura Y., Kaneko T., Kato I., Asamizu E., Tabata S.;			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=COLUMBIA;			
RX	MEDLINE=20277480; PubMed=10819329;			
RA	Nakamura Y.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and PAC clones.";			
RL	DNA Res. 7:131-135(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COL-0; TISSUE=GREEN VEGETATIVE TISSUE;			
RA	Weigel R.R., Pfitzner A.J.P., Pfitzner U.M.;			
RT	"NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins from Arabidopsis that interact in vivo with NPR1/NIM1, a key regulator of systemic acquired resistance in plants.";			
RT	of submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
RI	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AB023041; BAB01050.1; ..			
DR	EMBL: AJ250185; CAC19845.1; ..			
SO	SEQUENCE 122 AA; 13508 MW; 24FD70920CE1C152 CRC64;			

ALIGNMENTS

RESULT 1

ID	Q9LJA3	PRELIMINARY:	PRT:	122 AA.
AC	Q9LJA3			
DT	01-OCT-2000 (TRENBLrel. 15, Created)			
DT	01-OCT-2000 (TRENBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TRENBLrel. 15, Last annotation update)			
DE	Genomic DNA, chromosome 3, P1 clone: MPE11 (NIMIN-2 protein).			
GN	NIMIN-2.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
CX	NCBI TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=COLUMBIA;			
RA	Sato S., Nakamura Y., Kaneko T., Kato I., Asamizu E., Tabata S.;			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=COLUMBIA;			
RX	MEDLINE=20277480; PubMed=10819329;			
RA	Nakamura Y.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and PAC clones.";			
RL	DNA Res. 7:131-135(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COL-0; TISSUE=GREEN VEGETATIVE TISSUE;			
RA	Weigel R.R., Pfitzner A.J.P., Pfitzner U.M.;			
RT	"NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins from Arabidopsis that interact in vivo with NPR1/NIM1, a key regulator of systemic acquired resistance in plants.";			
RT	of submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
RI	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AB023041; BAB01050.1; ..			
DR	EMBL: AJ250185; CAC19845.1; ..			
SO	SEQUENCE 122 AA; 13508 MW; 24FD70920CE1C152 CRC64;			


```
DR HSP; P00918; lbv3.
DR InterPro; IPR001143; Euk_COanhd.
DR Pfam; PF00194; Carb_anhydase; 1.
DR ProDom; PD000865; Euk_COanhd; 1.
DR SEQUENCE 2254 AA; 248158 MW; 64E80CE274731FF3 CRC64;
Query Match 12.8%; Score 80.5; DB 10; Length 2254;
Best Local Similarity 21.7%; Pred. No. 85;
Matches 25; Conservative 25; Mismatches 34; Indels 31; Gaps 4;
QY 4 LKKEERV-----EEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFF 42
DB 232 NMKKDKIGLTGRTTYTRSLAASIPASVEQETPGLRRSSRGTPSTKVTTPASATRKSE-- 289
QY 43 KILRRVHVATRTVAKVNGGVAEGELPSKKRKRSQ---NLGLRNSLDC-NGVRDGE 93
DB 290 ----RLAPSPASVSKSGGIVKNSTPSSLRRNRGKTEVSLQSSKSGSDNSIRKGD 340
RESULT 5
Q9X5B3 PRELIMINARY: PRT: 753 AA.
ID Q9X5B3
AC Q9X5B3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Lactoferrin-binding protein.
SN LBPB.
SC Neisseria meningitidis.
OS Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
DX [1]
RX SEQUENCE FROM N.A.
RP STRAIN=M990;
RC MEDLINE=99250255; PubMed=12231574;
RA Pettersson A., van der Breen C., Joosten V., Hendriksen J.,
RA Tomassen J.;
RT *Sequence variability of the meningococcal lactoferrin-binding protein
RT LbpB";
RL Gene 231:105-110(1999).
RL ENBL: AF123381; AAD31769.1; ..
DR InterPro; IPR001677; transferrin_bind.
DR Pfam; PF01298; Lipoprotein_5; 1.
DR SEQUENCE 753 AA; 82180 MW; CE2670AB0389AALC CRC64;
Query Match 12.8%; Score 80; DB 2; Length 753;
Best Local Similarity 27.8%; Pred. No. 25;
Matches 25; Conservative 18; Mismatches 33; Indels 14; Gaps 3;
Y 6 KKEERVEEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFFKILRRVHVATRTVAKVNGGVAEG 65
b 496 EEDTSEEDNGEDEATAEETEVEDEAEVEE-----PEKSPAGNGC-SGS 545
Y 66 ELPSKKRKRSQNLGLRNSLDCNGVRDGEFD 95
b 546 ILPALEASKRGRI-----DPLKGIKRIATETD 571
RESULT 6
Q9M659 PRELIMINARY: PRT: 2001 AA.
ID Q9M659;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MOM.
SC Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
DX [1]
RX SEQUENCE FROM N.A.
RP STRAIN=CV. ZHUERICR;
RC MEDLINE=20279299; PubMed=10821279;
RA Amodeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
RT "Disruption of the plant gene MOM releases transcriptional silencing
RT of methylated genes";
RI Nature 405:203-206(2000).
RL EMBL: AF213627; AAF73380.1; ..
SQ SEQUENCE 2001 AA; 218586 MW; F6C5DAE897F2FE61 CRC64;
Query Match 12.8%; Score 79.5; DB 10; Length 2001;
Best Local Similarity 21.9%; Pred. No. 92;
Matches 25; Conservative 24; Mismatches 34; Indels 31; Gaps 4;
QY 5 LKKEERV-----EEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFF 43
DB 1 MKKDEKIGLTGRTTYTRSLAASIPASVEQETPGLRRSSRGTPSTKVTTPASATRKSE--- 57
QY 44 ILRRVHVATRTVAKVNGGVAEGELPSKKRKRSQ---NLGLRNSLDC-NGVRDGE 93
DB 58 ----RLAPSPASVSKSGGIVKNSTPSSLRRNRGKTEVSLQSSKSGSDNSIRKGD 108
RESULT 7
Q9M658 PRELIMINARY: PRT: 2001 AA.
ID Q9M658
AC Q9M658;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MOM.
SC Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
DX [1]
RX SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMB-A;
RC MEDLINE=20279299; PubMed=10821279;
RA Amodeo P., Habu Y., Afsar K., Mittelsten Scheid O., Paszkowski J.;
RT "Disruption of the plant gene MOM releases transcriptional silencing
RT of methylated genes";
RI Nature 405:203-206(2000).
RL EMBL: AF213628; AAF73381.1; ..
SQ SEQUENCE 2001 AA; 218569 MW; B9C85D4E62704441 CRC64;
Query Match 12.8%; Score 79.5; DB 10; Length 2001;
Best Local Similarity 21.9%; Pred. No. 92;
Matches 25; Conservative 24; Mismatches 34; Indels 31; Gaps 4;
QY 5 LKKEERV-----EEDNGKSDGNRGKPSCTEVVKTIVIEEVDFFF 43
DB 1 MKKDEKIGLTGRTTYTRSLAASIPASVEQETPGLRRSSRGTPSTKVTTPASATRKSE--- 57
QY 44 ILRRVHVATRTVAKVNGGVAEGELPSKKRKRSQ---NLGLRNSLDC-NGVRDGE 93
DB 58 ----RLAPSPASVSKSGGIVKNSTPSSLRRNRGKTEVSLQSSKSGSDNSIRKGD 108
RESULT 8
Q9K2A4 PRELIMINARY: PRT: 265 AA.
ID Q9K2A4
AC Q9K2A4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 18, Last annotation update)
DE Hypothetical protein CP0291.
GN CP0291.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83558;
DX [1]
RX SEQUENCE FROM N.A.
```

```
RC STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RA Read I.D., Brunham R.C., Sten C., Gill S.R., Heidelberg J.P.,
RA White O., Hickey E.K., Petersen J., Otterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RA pneumoniae AR39.";
RI Nucleic Acids Res. 28:1397-1406(2000);
RL EMBL: AE002191; AAF38248.1;
DR TIGR: CPC29;
KW Hypothetical protein.
W SEQUENCE 265 AA; 29523 MW; FA9B4AFBECF30FE CRC64;
30

Query Match 12.6%; Score 78.5; DB 16; Length 265;
Best Local Similarity 27.4%; Pred. No. 10;
Matches 40; Conservative 18; Mismatches 37; Indels 51; Gaps 7;

DY 13 EDNGKSGNRCK-----PSTEVRTVTEEE---VDEFFKTLR-RVHVATRTVAKVNSGV 52
db EDLGVKVGRTEDPVPVTPTEIVOLIPDEELSTVDEALQGRSLYAYKSVK----- 159
DY 63 ABGEPSKKRKRKRSQGLRNSL-----CNGVR-----DGEFD 95
db 160 -----PMQDLALVGFGLRDSAGLNFVRLANGVQNHYPHTKVKLYLAKNLAQWVDEIS 214
DY 96 EINRVGLQIGLGNCKPEPVSUS 121
db 215 EEKQQLRALGSE-----PKIESISIT 236

RESULT 9
19-JF8 PRELIMINARY: PRT; 252 AA.
AC Q9FJF8;
T 01-MAR-2001 (TREMBlrel. 16; Created)
T 01-MAR-2001 (TREMBlrel. 16; Last sequence update)
T 01-DEC-2001 (TREMBlrel. 19; Last annotation update)
E Genomic DNA, Chromosome 5, pl clone:MCM23 (Hypothetical 28.5 kDa protein).
S Arabidopsis thaliana (Mouse-ear cross).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
C eurosids II; Brassicales; Brassicaceae; Arabidopsis.
X NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.
X MEDLINE-99087469; PubMed-9872454;
X Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
X Tabata S.;
A "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
T Sequence features of the regions of 1,013,767 bp covered by sixteen
T physically assigned pl and TAC clones.";
L DNA Res. 5:297-308(1998).
P [2]
N P SEQUENCE FROM N.A.
A Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Cale J.M.,
A Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
A Tang C., Toriumi M., Yandura Y., Yu G., Yu S., Bowser J.,
A Kamiya P., Chen H., Cheuk R., Hayashizaki Y., Ishida S., Jones T.,
A Karlin-Neumann G., Kawai J., Kim C., Koesoma E., Lur R.,
A Lit J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
A Sakurai T., Satou M., Seki M., Shinozaki K., Southwick A., Tracy S.E.,
A Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
T "Full length cDNA of gene MCM23.14/AF591604 (GI:9757896).";
L Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
L EMBL: AB015473; BAB98403.1;
R EMBL: AY050965; AAK93642.1;
W Hypothetical protein.
Q SEQUENCE 252 AA; 28549 MW; 78DF092FA6C60D7F CRC64;
```

```
Query Match 12.5%; Score 78; DB 10; Length 252;
Best Local Similarity 39.3%; Pred. No. 10;
Matches 24; Conservative 10; Mismatches 17; Indels 10; Gaps 3;

CY 26 EVVRTVTEVEDEFFKIL-----RVHVATRTVAKVNGVAEGSLPSKKRKRKSNLGLR 81
db 131 EVIEF---KSLDEAFKLLIKQPGARLHVFSFDLQVNGEGYEG-LSSLKRKESRYVGLR 186
CY 52 N 62
db 187 D 187

RESULT 10
Q9TIF8 PRELIMINARY: PRT; 302 AA.
AC Q9TIF8;
DT 01-MAY-2000 (TREMBlrel. 13; Created)
DT 01-MAY-2000 (TREMBlrel. 13; Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19; Last annotation update)
DE Hypothetical 32.9 kDa protein.
OS Lactobacillus bacteriophage phi adh.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=12417;
RN 1;
RP SEQUENCE FROM N.A.
RA Altermann E.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN 2;
RP SEQUENCE FROM N.A.
RX MEDLINE-99384014; PubMed=10452953;
RA Altermann E., Klein J., Heinrich B.;
RT "Primary structure and features of the genome of the Lactobacillus
RT gasserii temperate bacteriophage phi-adh.";
RI Gene 236:333-346(1999).
RN 3;
RP SEQUENCE FROM N.A.
RX MEDLINE-95138034; PubMed=7836307;
RA Henrich B., Binshofer B., Blaesi J.;
RT "Primary structure and functional analysis of the lysis genes of
RT Lactobacillus gasserii bacteriophage phi-adh.";
RN 4;
RP SEQUENCE FROM N.A.
RX MEDLINE-93231538; PubMed=8472961;
RA Tremaux C., de Antoni G., Raya R., Klaenhammer T.;
RT "Genetic organization and sequence of the region encoding integrative
RT functions from Lactobacillus gasserii temperate bacteriophage phi-
RT adh.";
RI Gene 126:61-66(1993).
RN 5;
RP SEQUENCE FROM N.A.
RA Kage G., Altermann E., Klein J., Heinrich B.;
RT "Structure of a genome region of the Lactobacillus gasserii temperate
RT phage phi adh covering a repressor gene and cognate promoters.";
RI Gene 210:67-70(1998);
DR EMBL: AJ131519; CAB52529.1;
KW Hypothetical protein.
SO SEQUENCE 302 AA; 32916 MW; 6C953088D1A83109 CRC64;

Query Match 12.4%; Score 77.5; DB 9; Length 302;
Best Local Similarity 24.8%; Pred. No. 15;
Matches 32; Conservative 26; Mismatches 38; Indels 33; Gaps 5;

CY 1 MNSLSKKEEVEEDNCKSGNRKSTEVRTVTEVEDEFFKILRRVHVATRTVAKVNG 50
db 63 LGUTLKKQOSSLERN-KSEINSLKEAQSKVDKSTEGRNEYERYSKETATERNVATTT 121
CY 61 GVAEGELPSKKRKRKSNLGRNSLDC--NGV-----RDGEFDE 96
db 122 RIAT---LSCQQDRA-----RNSLDYKSGLASAQAALKISESSNAVYGRLEAGKHEE 173
CY 97 INEVGIQGL 105
```

```

RT glycosylatase and flanking genes in the cottontail herpesvirus
KI genome."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L33071; AAA91081.1; -
SQ SEQUENCE 325 AA; 36323 MW; DD9A987D4ABB1FE8 CRC64;

Query Match 12.1%; Score 75.5; DB 12; Length 325;
Host Local Similarity 30.4%; Pred. No. 25;
Matches 24; Conservative 17; Mismatches 25; Indels 13; Gaps
QY 9 ERVEEDNG--KSDGNRGK-----PS-TEVVRVTVEEYDEFFKILRRVHVATRTVAK 57
DB 167 ESMSENEGSTRKSRSGKRGCHGLRHPDPSIVEISSSESEDEVERF--FERRAEAVLRZYKK 244
QY 56 VNGVGAEGELPSKKRRKSQ 76
DB 225 KSGDSGDDLPSKLQQOAR 243

RESULT 13
Q9R3S1 PRELIMINARY; PRT: 364 AA.
AC Q9R3S1
CT 01-NOV-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane lipoprotein.
CC Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
CC Actinobacillus.
GX Actinobacillus.
OC NCBI_TaxID=715;
KN K1
KN K1
PP SEQUENCE FROM N.A.
RC STRAIN=PMOEX AND SL421;
RX MEDLINE=98125231; PubMed=9466755;
RA Gram T.; Ahrens P.;
KA "Improved diagnostic PCR assay for Actinobacillus pleuropneumoniae
KT based on the nucleotide sequence of an outer membrane lipoprotein."
KL J. Clin. Microbiol. 36:443-448(1998).
RL EMBL: U86681; AAD20614.1; -
RL EMBL: U86677; AAD20610.1; -
DR InterPc: IPR005014; lipoprotein_14.
DR Pfam: PF03346; lipoprotein_14; 1.
KW lipoprotein.
KW SEQUENCE 364 AA; 39575 MW; 4A82E539652485A9 CRC64;

```

Query Match	12.0%	Score 75;	DB 2;	Length 364;
Best Local Similarity	24.6%	Pred. No. 32;		
Matches	31;	Conservative	21;	Mismatches 58; Indels 16; Gaps
QY	5	KKEERVEEENKSDGNCK-----PSTEVKVTVEEEDFFKILRRHVAVR	53	
Db	109	KDQKLEEPKPKNSAEILKELGDKDKTGIITSDVVLNLTIDE-QENIQRLSESDIVR	167	
QY	54	TVAKNGGVAESELPKRRKRQNLGRNSLDCNGVRDGEFDEINRVLOG---IGLDELN	110	
Db	158	NCLKIINTIPNODITRLKDSFGRLGYYGYGMOLNVRGEGYGINNVLVGHYLLSKDES	227	
QY	111	CPPEPD	116	
Db	228	TKTAPN	233	
RESULT 14				
Q9R3V2		PRELIMINARY;	PRT;	365 AA.
ID	Q9R3V2			
AC	Q9R3V2;			
DT	01-MAY-2000 (TRENBLrel. 13, Created)			
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)			
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)			
DE	Outer membrane lipoprotein precursor.			
GN	OMA.			
OS	Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).			

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Actinobacillus.
NCBI_TaxID=715;
[1]
SEQUENCE FROM N.A.
RC STRAIN=WF83;
MEDLINE=98026959; PubMed=9809431;
RA Ito H., Osaki M., Uchida I., Ohya T., Sekizaki T.;
RT "Demonstration of the third antigenically distinct outer membrane
AT lipoprotein (OmlA) in Actinobacillus pleuropneumoniae serotype 7";
RE PEMS Microbiol. Lett. 157:303-308(1998).
[2]
RC SEQUENCE FROM N.A.
RC STRAIN=WF83;
RC STRAIN=WF83;
RX MEDLINE=96376964; PubMed=8782213;
RA Chevallier B.;
RT "Palindromic structure of the lqh3 locus control region";
RL Nat. Genet. 14:15-16(1995).
[3]
RC SEQUENCE FROM N.A.
RC STRAIN=WF83;
RC MEDLINE=98126231; PubMed=9466755;
RA Gram T., Ahrens P.;
RT "Improved diagnostic PCR assay for Actinobacillus pleuropneumoniae
RI based on the nucleotide sequence of an outer membrane lipoprotein";
RJ J. Clin. Microbiol. 36:443-448(1998).
[4]
RC EMBL; AB007579; BA334652.1;
RC EMBL; U86682; AAC0615.1;
RC InterPro: IPR005014; Lipoprotein_14.
RC Pfam: PF03346; Lipoprotein_14;
RC Lipoprotein; Signal; 19
RC SIGNAL
RC SEQUENCE 365 AA; 39708 MW; 3DFA543CC115C5F3 CRC64;

Query Match 12.0%; Score 75; DB 2; Length 365;
Best Local Similarity 24.6%; Pred. No. 32;
Matches 31; Conservative 21; Mismatches 58; Indels 16; Gaps 3;
2Y 6 KKEERVEONGKSDGNRGK-----PSTEVRVTVTEEEVDEFFKILRRVHVATR 53
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 109 KDDQKLEEPKNSAEILKELGDKIKTKGIIITRSDVNLTLDE-QENIQRLSESDIVR 167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
2Y 54 TVAKYNGVAGGELPSKKRKSQNLGRSLDCNGVRDGEFDEINRVGLQG--LGLDLN 110
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 168 NDLKIITNIPNQDITRLKDSIRGLGYYGMQLNQVREGERYGINNVDLVGHYLSMDPS 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
2Y 111 CKPEPD 116
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 228 TKTAPN 233

RESULT 15
2YFCG2
ID Q9RCG2 PRELIMINARY; PRT: 375 AA.
AC Q9RCG2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane lipoprotein.
OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=715;
OX [1]
RX SEQUENCE FROM N.A.
RC STRAIN=M62;
RC MEDLINE=98126231; PubMed=9466755;
RA Gram T., Ahrens P.;
RT "Improved diagnostic PCR assay for Actinobacillus pleuropneumoniae
RI based on the nucleotide sequence of an outer membrane lipoprotein";
RJ J. Clin. Microbiol. 36:443-448(1998).
[4]
RC EMBL; U86678; AAC00611.1;
RC InterPro: IPR005014; Lipoprotein_14.
RC

DR Pfam: PF03346; Lipoprotein_14; 1.
KW Lipoprotein.
SQ SEQUENCE 375 AA; 40840 MW; 67BA62BE58E01A02 CRC64;
Query Match 12.0%; Score 75; DB 2; Length 375;
Best Local Similarity 24.6%; Pred. No. 33;
Matches 31; Conservative 21; Mismatches 58; Indels 16; Gaps 3;
QY 6 KKEERVEONGKSDGNRGK-----PSTEVRVTVTEEEVDEFFKILRRVHVATR 53
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 120 KDDQKLEEPKNSAEILKELGDKIKTKGIIITRSDVNLTLDE-QENIQRLSESDIVR 178
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 54 TVAKYNGVAGGELPSKKRKSQNLGRSLDCNGVRDGEFDEINRVGLQG--LGLDLN 110
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 179 NDLKIITNIPNQDITRLKDSIRGLGYYGMQLNQVREGERYGINNVDLVGHYLSMDPS 238
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 111 CKPEPD 116
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 239 TKTAPN 244

Search completed: December 4, 2002, 09:25:04
Job time : 94 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Inc.

DN nucleic - nucleic search, using sw model

un on: December 4, 2002, 07:05:01 : Search time 3076 Seconds
(without alignments)
16094.125 Million cell updates/sec

title: US-09-733-685-3

effect score: 1700

sequence: 1 tgggtttttatggataac.....tttacgaactgaacatac 1700

coring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

searched: 2054640 seqs, 14551402878 residues

total number of hits satisfying chosen parameters: 4109280

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database :

GenEmbl:

1: gb_ba:

2: gb_ba:

3: gb_ba:

4: gb_ba:

5: gb_ba:

6: gb_ba:

7: gb_ba:

8: gb_ba:

9: gb_ba:

10: gb_ba:

11: gb_ba:

12: gb_ba:

13: gb_ba:

14: gb_ba:

15: gb_ba:

16: gb_ba:

17: gb_ba:

18: gb_ba:

19: gb_ba:

20: gb_ba:

21: gb_ba:

22: gb_ba:

23: gb_ba:

24: gb_ba:

25: gb_ba:

26: gb_ba:

27: gb_ba:

28: gb_ba:

29: gb_ba:

30: gb_ba:

31: gb_ba:

32: gb_ba:

33: gb_ba:

34: gb_ba:

35: gb_ba:

36: gb_ba:

37: gb_ba:

38: gb_ba:

39: gb_ba:

40: gb_ba:

41: gb_ba:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1700	100.0	83650	8	AB023041 Arabidops
C 2	576.2	33.9	92620	8	AB026636 Arabidops
C 3	540.2	31.8	83646	8	AB005248 Arabidops
C 4	534.6	31.4	94487	8	AC012394 Arabidops
C 5	534.6	31.4	100806	8	AC015450 Arabidops
C 6	499.8	29.4	104386	8	AT132A17 Arabidops
C 7	499.8	29.4	179771	8	ATCHRIV25 Arabidops
C 8	497	29.2	95519	8	AF071527 Arabidops
C 9	497	29.2	116448	8	AC005142 Arabidops
C 10	497	29.2	159629	8	ATCHRIV9 Arabidops
C 11	369	21.7	369	8	ATH250185 Arabidops
C 12	330.8	19.5	95190	8	AC007203 Arabidops
C 13	103.2	6.1	7218	6	I66494 Sequence 14 Arabidops
C 14	90.2	5.3	171763	2	AC115878 Mus muscu
C 15	89.2	5.2	280982	2	AC125184 Mus muscu
C 16	87.4	5.1	163362	2	AC101939 Mus muscu
C 17	87.4	5.1	217779	2	AL844515 Mus muscu
C 18	86.8	5.1	202872	2	AC016160 Homo sapi
C 19	85.6	5.1	166299	2	AC127244 Mus muscu
C 20	85.6	5.0	44735	9	AC092315 Homo sapi
C 21	85.6	5.0	48623	2	AC130469 Homo sapi
C 22	85	5.0	178036	10	AL672308 Mouse DNA
C 23	84.6	5.0	422	5	PD0M04 X93505 P. domesticu
C 24	84.6	5.0	51953	2	AC101966 Mus muscu
C 25	84.6	5.0	177648	10	AL732328 Mouse DNA
C 26	84.6	5.0	228944	2	AL844530 Mus muscu
C 27	84.4	5.0	222556	2	AC126558 Mus muscu
C 28	84	4.9	417	5	AF044763 Cecropis
C 29	84	4.9	202756	2	AC119804 Mus muscu
C 30	84	4.9	218898	2	AC114903 Mus muscu
C 31	83.6	4.9	162803	2	AC117818 Mus muscu
C 32	83.6	4.9	236538	2	AC102446 Mus muscu
C 33	83.4	4.9	239851	2	AC101712 Mus muscu
C 34	83.4	4.9	247340	2	AC121591 Mus muscu
C 35	83.2	4.9	64789	2	AC083839 Homo sapi
C 36	82.8	4.9	167364	10	AC121590 Mus muscu
C 37	82.8	4.9	220275	10	AL732404 Mouse DNA
C 38	82.6	4.9	48699	2	AC115677 Dictyoste
C 39	82.4	4.8	178413	2	AL805954 Mus muscu
C 40	82.4	4.8	203295	10	AL663049 Mouse DNA
C 41	82.2	4.8	204829	2	AC113318 Mus muscu
C 42	82	4.8	163132	2	AC097180 Rattus no
C 43	81.6	4.8	93661	9	AC107463 Homo sapi
C 44	81.6	4.8	106815	10	AL645809 Mouse DNA
C 45	81.6	4.8	120635	2	AC127463 Ornithorn

ALIGNMENTS

RESULT 1
AB023041/c
LOCUS AB023041 83650 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone: MPE11.
ACCESSION AB023041 BA000014
VERSION AB023041.1 GI:4220640
KEYWORDS
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui Pl clone:MPE11.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
REFERENCE 1 (sites)
AUTHORS Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.

Pred. No. is the number of results predicted by chance to have a

<p> TITLE Structural analysis of Arabidopsis thaliana chromosome 3. 1. JOURNAL MEDLINE REFERENCE AUTHORS JOURNAL COMMENT </p>	<p> Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones DNA Res. 7 (2), 131-135 (2000) 2 (bases 1 to 83650) Sato, S., Nakamura, Y., Kaneke, T., Kato, T., Asamizu, F. and Tabata, S. Direct Submission Submitted (01-FEB-1999), Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research: 1532-3, Yana, Kisarazu, Chiba 252-0612, Japan (E-mail: ynakam@kazusa.or.jp, Tel: 81-438-52-3335, Fax: 81-438-52-3334) Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd-graph.cgi?c=MPE11 Genes with similarity to proteins in the databases are described as 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Graal-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-061.mit.edu/GENSCAN.html), NetGene2 (S.M. Beagard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlins.zool.tastata.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K9122 and the 3' clone is M2114. Location/Qualifiers 1..83650 /organism="Arabidopsis thaliana" /strain="Colombia" /db_xref="taxon:3702" /chromosome="3" /clone="MPE11" /clone_lib="Mitsui PI" complement(542..764) /note="CDS is reported in Acc# AP000599 contains similarity to CHP-rich zinc finger protein gene_id:K9122.5" /number=1 /evidence=not_experimental join(1765..2463,2548..2739,2662..2649,3037..3107, 3190..3273,3410..3514) /note="gb:AA055139.1 gene_id:MPE11.1" /codon_start=1 /evidence=not_experimental /product="dihydrolipease de S-acetyltransferase" /protein_id="BAB01047.1" /db_xref="GI:9279589" /translation="MTVRSKTRIEFMPALSTMIIEGKIVSKIKTESKIAKGSVVVV ESDKADMDVTFYDGYLAAIVVGGETAVPGAAIGLLAEFABIEEAKSKAASKSSSS VAERAVPSPPVTSAPATAVPGAVTAVSGCAKTVATPKAKAKQKAVGIESVAG TGPRGRITASDVERAAGIAKSKSIAPPPPPPTAKATTNLPPLPUSSIVPTTA MQSAVKNMIESLVSPFRVGPVNTDADLALYIEKVKPKGYMTTA:LAKAAGMALAJH PVNAACKDGKGSFYSNSINIAVAINGSLITPVIQADKLDLYLQSKKELVGKA RSKLOPHVYNSGTFTLSNLGMPGVDRFDAILPPGOGAIMAVGASKPTVYADKGFPS VKNTMLVNTAHRIVYCADLAAAFITQFAKTIENPDSITL" complement(4594..5106) /note="gene_id:MPE11.2 unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01048.1" /db_xref="GI:9279590" /translation="MEGLLEELRTQSVGNKRVKRDGLDLSPPVKRLRDLFDSDSL DPVSDIDSVNKSFTENSLTAAISGTEUPOGLYFRASDDGLGPPPTPTFCTLI PPSCEETVELVRASSDSSEYGE:CGFEDHVTETFCPCDLSDDGLFEYFDGCLSDGLF </p>	<p> CDS SRRPFLPAE" complement(join(10090..10408,10649..10804,10916..10962)) /note="gene_id:MPE11.3 similar to unknown protein spIP42744" /codon_start=1 /evidence=not_experimental /protein_id="BAB01049.1" /db_xref="GI:9279591" /translation="MNEPKARYDRLMYTIQOTLEASICLLNCGPIGSKALKNLVLS GVGSTIVIGSKVLIGDKQFHRHAEIKFSEISGFRDENTVFQRSEQSVFQRQ, EDRNAGQ:VRPMQIARRIWAISGRWLSADNTITYGKEARAITDPPGRVLARLT, VDRHLDPSRLQDI" complement(11450..11818) /note="gene_id:MPE11.4 unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01050.1" /db_xref="GI:9279592" /translation="MNSLKKEERVEEDNGSKDGNRKPSTEVVTVTEVEEDVEFKI LRRHVARTVAKVGVGAEGLPSKRRKRSNGLRNSLDGNGVDFGDFEINRVGL QGLGLDIACKPPEPDSVLSL" 1698..17979 /note="gene_id:MPE11.5" /codon_start=1 /evidence=not_experimental /product="AP2 domain transcription factor-like protein" /protein_id="BAB01051.1" /db_xref="GI:9279593" /translation="MAERKKRSIOTNPKPKPKKPPOLNHLPLGLSGLTKMKRJR FVNDPYATYSSSEEBESQRRKYVCEILDLPFAQAATOAESESSYQOESNNNGVK TKISCSKVLRSKASPVVGRSTVSPVGVQRKWKAAEIRHPITKVRWLGTY EILEQAADATKLEFDALAAATSAASVLSNESMSISAGSSIDJLKLVDSTLD QQAGESKASPDFADLIQPEMGCFIDDSFIPNACELDFLTENNQMDDYCGID DLDI:GLECDGSPDPYDSDVEIDLGLIGTIIDYKAFVDHIATTIPTPLNIACP" join(21893..22063,22330..22386,22578..22783,22894..23227, 23311..23382,23473..23613) /note="gb:AAAF23821.1 gene_id:MPE11.6" /codon_start=1 /evidence=not_experimental /product="homocysteine S-methyltransferase AtHMT-1" /protein_id="BAB01052.1" /db_xref="GI:9279594" /translation="MVLEKKSALLEDLKKCGCAVVGGFATQLEIHAALNDPLKS AVS:KNPELIRKRMHELEAGADIVTSSYQATIPGLSRGLSEESLQKSVEL AVEARDVKEVKVSGSHSYNRYALVAASIGSYGAYLADGEYSYGHYGNVSLDKLDF HRRRLQVLEAGDOLLAFETIPNLEAQACVELLEEKVQIPAWICFTSVGDEKAPSG ESFECELPNKSNNIYAVGINCAPQFTENIRKPAKTKKAIIVVYNSGVEWDGKA KQNLPSQCFGDDFEFMFATKWRDLGAKLLIGCCRTTTPSTINAIRDLKRR" complement(27277..27552) /note="gene_id:MPE11.7 unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREKSSLTITLLMILLSYMLHFCVYSRYCAIMHFPETP ASGRKEEDLMKKYFGAGKFPFVDFGVGIGSEKSRIVPSCPDPLN" 30765..31883 /note="gb:AAAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKEREKRSVLDKYKRSPLQCEASLAKPSEKQKWEKAEARC? VKMEHFNHILLICSSYNGCRPYMCDTSHRNSCFDFPKASKPKPSLLRBEES NPTMEDVDSSTAVNLGEAASEITVVDLSDGGEVEEVEEVEEVEEVEEVEE ITTEDEKKNPKQKLCPLCRKIVKVVYKAAKCFMKNKRSCTCETCDFSGSYDLIR KHARLLHGPCVRPSEADPQRQRWRRLERQSLGLLQSLQSSFGGDEISNDGDFLAD </p>
<p> FEATURES source exon: </p>	<p> CDS join(1765..2463,2548..2739,2662..2649,3037..3107, 3190..3273,3410..3514) /note="gb:AA055139.1 gene_id:MPE11.1" /codon_start=1 /evidence=not_experimental /product="dihydrolipease de S-acetyltransferase" /protein_id="BAB01047.1" /db_xref="GI:9279589" /translation="MTVRSKTRIEFMPALSTMIIEGKIVSKIKTESKIAKGSVVVV ESDKADMDVTFYDGYLAAIVVGGETAVPGAAIGLLAEFABIEEAKSKAASKSSSS VAERAVPSPPVTSAPATAVPGAVTAVSGCAKTVATPKAKAKQKAVGIESVAG TGPRGRITASDVERAAGIAKSKSIAPPPPPPTAKATTNLPPLPUSSIVPTTA MQSAVKNMIESLVSPFRVGPVNTDADLALYIEKVKPKGYMTTA:LAKAAGMALAJH PVNAACKDGKGSFYSNSINIAVAINGSLITPVIQADKLDLYLQSKKELVGKA RSKLOPHVYNSGTFTLSNLGMPGVDRFDAILPPGOGAIMAVGASKPTVYADKGFPS VKNTMLVNTAHRIVYCADLAAAFITQFAKTIENPDSITL" complement(4594..5106) /note="gene_id:MPE11.2 unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01048.1" /db_xref="GI:9279590" /translation="MEGLLEELRTQSVGNKRVKRDGLDLSPPVKRLRDLFDSDSL DPVSDIDSVNKSFTENSLTAAISGTEUPOGLYFRASDDGLGPPPTPTFCTLI PPSCEETVELVRASSDSSEYGE:CGFEDHVTETFCPCDLSDDGLFEYFDGCLSDGLF </p>	<p> CDS SRRPFLPAE" complement(join(10090..10408,10649..10804,10916..10962)) /note="gene_id:MPE11.3 similar to unknown protein spIP42744" /codon_start=1 /evidence=not_experimental /protein_id="BAB01049.1" /db_xref="GI:9279591" /translation="MNEPKARYDRLMYTIQOTLEASICLLNCGPIGSKALKNLVLS GVGSTIVIGSKVLIGDKQFHRHAEIKFSEISGFRDENTVFQRSEQSVFQRQ, EDRNAGQ:VRPMQIARRIWAISGRWLSADNTITYGKEARAITDPPGRVLARLT, VDRHLDPSRLQDI" complement(11450..11818) /note="gene_id:MPE11.4 unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01050.1" /db_xref="GI:9279592" /translation="MNSLKKEERVEEDNGSKDGNRKPSTEVVTVTEVEEDVEFKI LRRHVARTVAKVGVGAEGLPSKRRKRSNGLRNSLDGNGVDFGDFEINRVGL QGLGLDIACKPPEPDSVLSL" 1698..17979 /note="gene_id:MPE11.5" /codon_start=1 /evidence=not_experimental /product="AP2 domain transcription factor-like protein" /protein_id="BAB01051.1" /db_xref="GI:9279593" /translation="MAERKKRSIOTNPKPKPKKPPOLNHLPLGLSGLTKMKRJR FVNDPYATYSSSEEBESQRRKYVCEILDLPFAQAATOAESESSYQOESNNNGVK TKISCSKVLRSKASPVVGRSTVSPVGVQRKWKAAEIRHPITKVRWLGTY EILEQAADATKLEFDALAAATSAASVLSNESMSISAGSSIDJLKLVDSTLD QQAGESKASPDFADLIQPEMGCFIDDSFIPNACELDFLTENNQMDDYCGID DLDI:GLECDGSPDPYDSDVEIDLGLIGTIIDYKAFVDHIATTIPTPLNIACP" join(21893..22063,22330..22386,22578..22783,22894..23227, 23311..23382,23473..23613) /note="gb:AAAF23821.1 gene_id:MPE11.6" /codon_start=1 /evidence=not_experimental /product="homocysteine S-methyltransferase AtHMT-1" /protein_id="BAB01052.1" /db_xref="GI:9279594" /translation="MVLEKKSALLEDLKKCGCAVVGGFATQLEIHAALNDPLKS AVS:KNPELIRKRMHELEAGADIVTSSYQATIPGLSRGLSEESLQKSVEL AVEARDVKEVKVSGSHSYNRYALVAASIGSYGAYLADGEYSYGHYGNVSLDKLDF HRRRLQVLEAGDOLLAFETIPNLEAQACVELLEEKVQIPAWICFTSVGDEKAPSG ESFECELPNKSNNIYAVGINCAPQFTENIRKPAKTKKAIIVVYNSGVEWDGKA KQNLPSQCFGDDFEFMFATKWRDLGAKLLIGCCRTTTPSTINAIRDLKRR" complement(27277..27552) /note="gene_id:MPE11.7 unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01053.1" /db_xref="GI:9279595" /translation="MTHAREKSSLTITLLMILLSYMLHFCVYSRYCAIMHFPETP ASGRKEEDLMKKYFGAGKFPFVDFGVGIGSEKSRIVPSCPDPLN" 30765..31883 /note="gb:AAAF16548.1 gene_id:MPE11.8 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAB01054.1" /db_xref="GI:9279596" /translation="MPKEREKRSVLDKYKRSPLQCEASLAKPSEKQKWEKAEARC? VKMEHFNHILLICSSYNGCRPYMCDTSHRNSCFDFPKASKPKPSLLRBEES NPTMEDVDSSTAVNLGEAASEITVVDLSDGGEVEEVEEVEEVEEVEEVEE ITTEDEKKNPKQKLCPLCRKIVKVVYKAAKCFMKNKRSCTCETCDFSGSYDLIR KHARLLHGPCVRPSEADPQRQRWRRLERQSLGLLQSLQSSFGGDEISNDGDFLAD </p>

CDs
TRLLIVYLIVYFPESSGSSSSWSSTSRKHTHSGRKSSRSPASLWSESYSZENTUT
SPROENQSSDEGVSTRRRSRRRTVVDDEDEEF
complement(join(32197..32430,32525..32676,32969..33332,
33464..33487))
/note="gene_id:MPell.9"

/codon_start=1
/evidence-not_experimental
/product="chloroplast 50S ribosomal protein L15"
/protein_id="BAB01055.1"
/db_xref="GI:9279597"
/translation="MAPLSLSSNPLSRHCYRLHLSSFKGNVSLGNPSSOILSL
KLNQTLKRNQOQFARPEVYVQIATANSSAVAPERLNLDPQPSRKKQKRGK
VSKGASCGMGKQSGRMGRFGGQALYRLRLKRLGIAGMSRSLPLKYP
INQIDTAGOEDGVDSLTKGKLINPNSGRERKLPLILIGELUSMLKLTFAKAF
STOAKRELEAGSGLTILPLGRKKVVPKSVAKNOARADEYFAKRRAAAEAAATSEPAS
A"

CDs
complement(join(33961..34021,34114..34226,34391..34482,
34602..34800))
/note="db|AAF26483.1
gene_id:MPell.10
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01056.1"
/db_xref="GI:9279598"

CDs
/translation="MKNVMLIDESNASYDLILWALENOKDITESSVYIFAKQPNS
FTFPLVSSVGFQIPIFYPSNSELRLACGRKNKRIALGILEKAKICLNHGKAE:
FDDGQKDLJRKIQEINILNIVSQOOLKQNTQDCSLVKKRQRKD"
join(36033..36259,36243..36283,36497..36565,36651..36707,
36797..36871)
/note="contains similarity to transcription-associated
zinc ribbon protein
gene_id:MPell.11"

CDs
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01057.1"
/db_xref="GI:9279599"
/translation="MEKRSSEFLFCNLGTMVLKSTKYAECPCPKTTRNAKCIUK
SIATVSAEDIRRELGLSLFCEKTOAEALPKIKACEKQHPPLVYTTTC:TSANEG
QTYTTCPCAHRETEG"
join(37285..37965,39579..38674)

Query Match 100.0% Score 1700; DB 8; Length 83650;
Best Local Similarity 100.0%; Pred. No. 2.1e-285;
Matches 1700; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 TGGGTTTATTGGTAACATGACAAATATTATTATTCATGAGTTTATTGGATAG 60
D 12680 TGGGTTTATTGGTAACATGACAAATATTATTATTTCATGAGTTTATTGGATAG 12621
Y 61 CATGACAAATATTAATATA:CAAGTGTAAATAACATGTTTGTTCATAAATACATGCAT 120
D 12620 CATGACAAATATTAATATA:CAAGTGTAAATAACATGTTTGTTCATAAATACATGCAT 12561
Y 121 TTAANATCAGACATTTGTTTAAATGAATCAATCTTATATATATATATCAAGACATTGAC 190
D 12560 TTAANATCAGACATTTGTTTAAATGAATCAATCTTATATATATCAAGACATTGAC 12501
Y 181 GGAATAATCAGGTAAAAAGAGAAATAAGAAATAGAGATAGAGAGATTTCATGGAAA 240
D 12500 GGAATAATCAGGTAAAAAGAGAAATAAGAAATAGAGATAGAGAGATTTCATGGAAA 12441
Y 241 AGAAGACAGACATGTAGGTGAACAAAATAAGAGATATGATGATAATTTATGAGAG 300
D 12440 AGAAGACAGACATGTAGGTGAACAAAATAAGAGATATGATGATAATTTATGAGAG 12381
Y 301 GTGCTGAAGATTATTTAGCAGGGGAGAGAGAAATAGAAAAGAAATGACATGCTGAAA 360
D 12380 GTGCTGAAGATTATTTAGCAGGGGAGAGAGAAATAGAAAAGAAATGACATGCTGAAA 12321
Y 361 TCTGAAGACAGATGAATTTGCTTAAAGATGAAGAGAGAAAGAGAAATCCATGCTTAAAGTC 420
D 12320 TCTGAAGACAGATGAATTTGCTTAAAGATGAAGAGAGAAAGAGAAATCCATGCTTAAAGTC 12261

```
1501 GGCCTTTGGCTGGCCGCTTACTATAATATTCAGAGTAGCTAIGGCCTGTATAAGGAGT 1560
|||||
11180 GGCCTTTGGCTGGCCGCTTACTATAATATTCAGAGTAGCTAIGGCCTGTATAAGGAGT 11221
|||||
1561 CCAAAAAGAGCGGCGCAATTGCGGCTTGGGCTTGGAACTTTGGAATTTGGAATTTG 1620
|||||
11120 CCAAAAAGAGCGGCGCAATTGCGGCTTGGGCTTGGAACTTTGGAATTTGGAATTTG 11061
|||||
1621 CATGTTTATCATTAATGCTGCGGACTTTGAGACGACGGCTTGTTCTTCTCCTG 1680
|||||
11060 CATGTTTATCATTAATGCTGCGGACTTTGAGACGACGGCTTGTTCTTCTCCTG 11061
|||||
1681 TTACGACTTACGAACATAT 1700
|||||
11000 TTACGACTTACGAACATAT 10981
|||||

RESULT 2
LOCUS AB026636 92620 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K14A17.
ACCESSION AB026636
VERSION AB026636.1 GI:4757392
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
P1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
2 (bases 1 to 92620)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0842, Japan (E-mail:ynakamura@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg-graph.cgi?c=K14A17
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown: protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli1.zozi.ilstate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNASCAN-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNASCAN-SE/).
This sequence may not be the entire insert of this clone, it may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MUI.5 and the 3' clone is MCE21.
Location/Qualifiers
1..92620
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="K14A17"
/clone_lib="Mitsui TAC"
complement(201..2180)
```

CDS

```
/note="gene_id:K14A17.1"
/codon_start=1
/evidence="not_experimental"
/product="salt-inducible protein-like"
/protein_id="BAA94973.1"
/db_xref="GI:7670019"
/translation="MRGFASSASRTATAAASKSLNASTSVNPKLSKTLNSSGKPTNP
LNQRYISQVIEPKDMFLINQBEFTTHRIGLARTFVIVLONQDPLHSEKRYLVSN:
DPVYAKDSKLSVGLNALFRGKPLLSMELKAEIRDSGYRISDELMCVLISWRGLG
AYINDFAOISFLGKMPSTRLYNAVIDALVKSNSLDLAYLKFQOMRSDGCKPQRTY
NILHGVCKGVYDEAIRLVKQMEQENPNVFTYITLIDFLGIAGRYDEALKLENM
RVKLNQPEARTIRTFVHGTFRCPLPCAKFEVLVGPMEKDSNLRQYGVDAVLYCSNNS
MAKETGPNLAKTIGERYIIPDSSTINAAKSLCKLGHDLVETCRIFDGFVRGVRKPGFN
YIVLVQALLNAORFSEGDRYLKQMGVDGLSSVYSYNAVIDCLOKARIEANAFLEIE
MQDRGISPNLVTNFTLSGYSVRGDKVKVHGVLEKLLVHGKFPDVTLSLINCJORA
KEIKDAFCFKEMLEWIEPNEIYINLIRSCSTGTDTRSVKLFARKKNGLSGPLY
AYNATIQSCFKMKRYKKAELKTNLRIGLKPDNFTYITLILKALSESRESEAREMEF
STEREGCVDSYTKRLVELDLRKSLGREIVSAS"
join(3906..4124,4238..4462,4536..4649)
/jote="emb|CAB82745.1
gene_id:K14A17.2
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAA94974.1"
/db_xref="GI:7670020"
/translation="MAKQDPTSSNTMLVGSSEKVRPOPLDJITGDTVKDAI3NKVF
KVKTLPLGLHNRRLVDPNDSPIVTKMKKVISKHQRWQYRGSDLDKRFIVKRSSTV
QKTRVEYFLKHNTRFESSCTIKGRFMKRACTIYVADSTKIIAQVYEGHERLVATI
YFNVAFTVITLIFIDLINMGTI"
complement(join(5031..5927,6011..6601,6984..7211))
/note="gene_id:K14A17.3"
/codon_start=1
/evidence="not_experimental"
/product="AMP-binding protein:"
/protein_id="BAA94975.1"
/db_xref="GI:7670021"
/translation="MAATKWRIDDLPKIPANYTALIPLWFLDRAAVVPTPKSVIHG
SREYTRQYDRCRLASALADRSIGPGSTVFVAILIAPNIPAMYEAHGVPMCGAVIN
CVNIRLNATVAFLSHSSOVSIMVDOERTFLAEDSLRLMEKAKSSKRPILLIVGD
LHCAPESLNRKSKAIEYEDFLATGDPNPQPPADEWQSIALGTSITRSPKGVV
LHGRGAYIMALNSPLIGMQDGAIVYLLPMPHCGWCFPWSLAVLSGTSICIRQVTA
KEVSMIAKYVTHFCAPAVNLNAIYNAPEKDTILPHTVMTAGAAPPSVFSM
NOKGPRVATYGLSTFCTGPTVCANKPEWSDLPETPAKLNARQVRYGIMOLDVIC
TQTKGVPVADGKTAGEIVFRGNMVKGLKPNKANKETFAGGHEHSGDIAVHKNYI
EIKORSKDYIISGGENISSVEVENYVYHHPAVLEASVYARPERHQSCEAFVILKSD
YEKDNQKLAQDLMKFCRKLPAIVVPKSVFPGLPKTAIGKIQKHILKRAKMGVP
PSRL"
complement(join(8463..8892,8977..9136,9230..9641))
/note="gene_id:K14A17.4"
/codon_start=1
/evidence="not_experimental"
/product="basic chitinase"
/protein_id="BAA94976.1"
/db_xref="GI:7670022"
/translation="MVSKPLFLSILLITVALVVFQGTILVNAEDSEPSSTRKPLKVI
KGLKLCQKGECKGWSEYCCNHTISDFETQYFENLFSKRNSPVAHVGFMVRSFIF
AAAEYQPLGFTAGKLGQMKVEAFLGHVSKTSCGTVGATGVLNGLCYKEMSP
DQYICDDYKLYLPTCTPGYSYHGRGALPYWNNYNGQGEALKVLLSHPELENNAT
LAFQALRWMTTPPKKHLPSSAHDVFGKWKPKTNDTAARTPGFGATINVDQDQICN
SCFDNDMMNIVSHYLYLDLIGVGRGEEAGPKLSCADQEPFSSSSAPSSGSSS"
join(13759..13822,14371..14445,14592..14707,14800..14888,
15014..15059,15238..15433,15520..15845,15928..17178,
17360..17713)
/note="gene_id:K14A17.5"
/codon_start=1
/evidence="not_experimental"
/product="transcription factor-like protein"
/protein_id="BAA94977.1"
/db_xref="GI:7670023"
/translation="MDQDGLGRIGISEIHGFHTLQDLVQTMLEAKSKRLPNPIHA
IUCGRILLDFLRKMLRNRKGNHKKDKGTGRTKAHEHLKVGNEIRHIVYAHGEDN
ITFVRRCYLLKARENIVLVHYRDTQEAATTSQDSISSPISVSQTTFFNRVAAEDID
```

CDS

```
TVVRNEDLSLDINLIDWJELVPTCLNNUSAPTVDNLISYETFEQLONRANGTAERENA
TVAGSLDALLNDGQSGSFGFRWNSEISESGSLEDSEFPFPMVPCDIEAPGVAF
HSHSNIPQVFNITDVSFAMVYSSKIKILVTGFLHDSYQHIFRSLNLYCVAGGCVFA
EYLOAGVYKIRIPPSIPGMVNLISADGKHPIPSQCFREHRAVPVLDKIVPECDNDSK
WEEFEQVRLSHLLTSSNKNVLSSKISPHNRDAUKLASKNHLNLLWAYLVASIO
KNKVSFDQAKHILFELSLNKLRLKMLMEKVLPGRNTLDYDSKGLGVHLCASLGYTWS
VOLFLSGLSLNFRKQGTALHRAAYVGRKNVVAALISAGARPNVDSFKDNLGGC
MAADLAQNGYDGLAAYLAERKLCVQAFQDMKTAGNITGDLEACKRMLNQGSLPDEEC
RYDFFRRIEAYRIQCFQFTWKIRREYLNMRQAIRIOAFRGICARROYKILWSV
GVLEKAVLRKQKRGKGFGLVAAEDSPGEAQEDFYKTSRQAEERLSRVVRQAM
PSKKAQODYRMKLTIDHEAOVNHITFLNLSFGKKNRR*
complement(1015,(18170..18172,18407..18490,18736..18807,
18977..19090,19168..19281,19387..19487,19590..19668,
19755..19805,19906..19969,20365..20431,20539..20669,
20916..21040,21126..21231,21349..21483,21624..21770,
21860..22171))
/note="gene_id:K14A17.6"
/pseudo
/codon_start=1
/evidence-not_experimental
/product="Lipoamide dehydrogenase"
join(29036..29206,29695..29826,29925..30009,30149..30312,
30830..31267)
/note="contains similarity to similar to ubiquitin
conjugating enzyme
gene_id:K14A17.7"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94978.1"
/db_xref="GI:7670024"
/translation="MADEBYNKNPAVKELIQEVKMQANSDQFMSLPLENIEHQ
FAIRPGDTEFGGIFHGRIQPADYPPKPPFMLTPNGRPETNKICLSINTEPE
HWQPSVSTALVAGIAPFTSPNGALGSVDYKDPERRTLAISKETPKYSIPEROK
IIDEHQYTSKATVYKPLPECSQAPSVSSEHQSVEQPEAITVVERSALTDCD
VDDQITFEAFVAVNTAASVVAAPALPVAVEVWVXASVSGEQRKARRAAGKPVDDRLEI
WAAGLTITAIWLLIKKFKSKNGYSTGPMDDGS"
join(32260..32364,33121..33412,33506..33750,33837..34103;
/note="golAD26962.1
gene_id:K14A17.8
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94979.1"
/db_xref="GI:7670025"
/translation="MKSSNIVFDRIKERQMLSPFKIFQRADLSQCKRALPYEFTS
KYSEKDFSKMVIKQAGXIGIVVEYSKNPTYYIETRGKDLFVSDNALQNEFIETH
RGMVPHVNIYEONGLEMKPRKFTQMPSPSGIKKEEGENSLIDYKKESESPSORA
EFLVRKKKCEDSKSKKMKTRNKVKKSKSKQVLDGVPEFKITRKSYIKFLATIK
HFVDDIKPKSKSIFITRHNGSKWKVCLVRLRITFISSGYSKLAHEFELMGDKITF
KLKIPFEVLLSKKRNKQDCMID"
complement(1013(34528..34725,35213..35322,35660..35865))
/note="gbIAAF26101.1
gene_id:K14A17.9
similar to unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94980.1"
/db_xref="GI:7670026"
/translation="MAESGGRIGVAVDFSDSKKALSWAIONVVVDGRLILITIAH
DMNVEGEMQLWETVCSGPIPMSEFSDAAVMKYALKPDAETLDIVNTAARKKITVY
MKIYWDGPREKICAAAEQIPLSLVMNGRCLOGLKRIMGVSYSVHVVNVVACPVTVK
AHI"
complement(1013(36309..36416,36556..36846,37250..37324,
37500..37634,38020..38145,38221..38374,38433..38573,
38878..38987,39134..39281,39474..39751,39993..40262))
/note="gene_id:K14A17.10
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA94981.1"
/db_xref="GI:7670027"
/translation="MDTNGASLIEIQDEEVEDPFLAFLDYARTVISPECDDEKEF
```

```
Query Match: 33.9% Score 576.2 DB 8: Length 92620;
Boot Local Similarity 99.5% Pred. No. 1.1e-90;
Matches 578; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TCGSGTTTATTGGATAACATGACAAAATATTTATTTTCATGAGTTTATTGCGATAG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71476 TGGGTTTTIATGGATAACATGACAAAATATTTATTTTCATGAGTTTATTG 71535
QY 61 CATGACAAAATATTAATATATCATAGTGTATTAACATGTTTGTCTTAAATACAT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71536 CATGACAAAATATTAATATATCATAGTGTATTAACATGTTTGTCTTAAAT 71595
QY 121 TTAATAATCAGACATTGTGTTTTTAAATCAATCTAATCTTTATATCAACAC 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71596 TTAATAATCAGACATTGTGTTTTTAAATCAATCTAATCTTTATATCAAC 71655
QY 181 SGNAATTCAGGTAAAGAGAGAAAATAAAGATGAGATACAGAGATTTCIATG 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71656 SGAAATTCAGGTAAAGAGAGAAAATAAAGATGAGATACAGAGATTTCIAT 71715
QY 241 AGAAGAGAGACATGTAGTGAACAAAATAAAGAGATATGATATATTTTATGAG 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71716 ACAAGAGAGACATGTAGTGAACAAAATAAAGAGATATGATATATTTTATG 71775
QY 301 GTGTCACAGATTATTTAGGAGAGGGAGAGAAAATAGAAAAGAAAATGACAT 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71776 GTGGTGAAGATTATTTAGGAGAGGGAGAGAAAATAGAAAAGAAAATGAC 71835
QY 361 TCTGAAGAAGATGAATTTGTTTAAAGATCAAGAGAGAAAGAACTCCATGGCT 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71836 TCTGAGAGAGATGAATTTGTTTAAAGATCAAGAGAGAAAGAACTCCATGGCT 71895
QY 421 TCGTAAAGAGATGAAAAAGAACAAAAGAGAGAGAAAGAAAGCGTAAATATA 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71896 TCGTAAAGAGATGAAAAAGAACAAAAGAGAGAGAAAGAAAGCGTAAATATA 71955
QY 481 CTAACATATTGCCAAAATTTCTGTAGCCGACAAATCTATTGTCACAGGTTATT 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 71956 CTAATATTGCCAAAATTTCTGTAGCCGACAAATCTATTGTCACAGGTTATT 72015
QY 541 TATTCITTTGAAGTCAAAAGTTATTTCTTACATATACTCTA 581
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72016 TATTCITTTGAAGTCAAAAGTTATTTCTTACATATACTCAA 72056
```

```
RESULT 3
AB005248/8
LOCUS 83646 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone MX110.
ACCESSION AB005248 BA000015
VERSION AB005248.1 GI:2264320
KEYWORDS
SOURCE
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui PI
clone:MX110.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (sites)
AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
Miyajima,N. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned pl clones
JOURNAL DNA Res. 4 (3), 215-230 (1997)
MEDLINE 97471969
REFERENCE 2 (bases 1 to 83646)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
```

COMMENT

tel:81-438-52-3935, Fax:81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/aga_graph.cgi?c=MX110
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Graft
 (Informatics Group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/Graft-1.3/>),
 GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),
 NetGen2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, <http://www.cbs.dtu.dk/services/NetGen2/>) and
 SplicePredictor (Volker Brendel, Stanford University,
<http://gremli.ni.zool.iastate.edu/cgi-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MX17 and the 3' clone is MR8-B.

FEATURES

source

1. .83646
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MX110"
 /clone_lib="W. Mitsui pl"
 join(1791..2912,3016..3318,3584..4560)
 /note="gene_id:MX110.1"
 /codon_start=1

CDS

/evidence=not_experimental
 /product="disease resistance protein-like"
 /protein_id="BAB09346.1"
 /db_xref="GI:9758812"

/translation="MTEKATDVSNMNNYSPSRDFDLGICQAKMKEMESLCLDS3
 EVRMIGSPGIGKT...ARVLYSOFSENFELSIEMGNIKELMTTRPCVSUEYSAK...Q
 LQKQFSLQINKHMDLHLHGVAGPRLNCKVLYLVDSDIOSIQ...DAKATKWFHG
 SRLIITQOKLLKANGINIKYVPPSYAYQMFCKARGQNPNDGFELAKVYI
 KLGLHPLGRVNGSHFRGSMRWNALPRUKIRIDASTISILKSTFALCEPCRL
 FLIIACLFNQNEVEVDYLAUSELDVROGFHLAEKSLINLKLEINCTIEMENL
 VQKGDIVRKHQKQICCPGRQFELIDARDICEYTDNINRNRYVGIFFLEVRNLSC
 LNTSERAFQMSMKLPRPHDPYDQSDOKLYLPQGLNLPQKRLIEMSRPMTCLFS
 NPTCYLYIRMKNSLQNLQWGNPLGNLKMDSLSEKHLKELPDLSTANLNYLIM
 SGCSILVPSPSIGLKLRKLMLSLRCQSKLEALPTNINLESLOVDLDDCLLIKKEPE
 LSNIKDLIKITATKXEVSTIKSNHLRKLKPSYSENKLEPHALDZTTLYINDPE
 MQPIQWYKXISHLOTGLGEGKRLVTIPQSDLSQLVYNGCESLERLNSFONHPE
 RFUWFNCFKLNEAREFTQTSHTAILPSREVPANFTRANGSSIMYNLNRRLSTI
 LRFKACVLLYKIKDNDKEAADRTVPIRENDKIGVDVPMWYRFHVPTILEDHL
 TFQIEADVTSNKLFSFDIKDNNQAVIKECGVLOZ"
 complement(join(5504..5620,5688..5736,5836..5905,
 5990..6122,6206..6250,6374..6688))
 /note="gene_id:MX110.2
 pir|IT03994
 similar to unknown protein"

CDS

/codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09347.1"
 /db_xref="GI:9758813"

/translation="MMASSLSFPCLSYKVPYIAATNSPATTIPYKXGOTRILH
 KIPSGLKMEVIEQPKSKSRPNPPVVFVSSVFAAKWAENKLPFFSSSGTDSYAVS
 LGGSEDEPLGTVACLTQASHDIADFTESNLGSSRPVVGHSFGGLVVOYLIANVN
 KRLSGTENAPTELVGAVNCSVPSPGNSGLVIRLFENKPAAPKVLTLAAKGFQKST
 PLCRETFFSQAMDEQLVKRYNAR"
 join(7788..7835,8043..8777,9032..9157,9205..9901,
 10085..12958)
 /note="gene_id:MX110.3"

CDS

/pseudo
 /codon_start=1

/evidence=not_experimental
 /product="non-LTR retroelement reverse transcriptase-like
 protein:"

CDS

complement(13525..13827)
 /note="gene_id:MX110.4
 pir|IT02618
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09348.1"
 /db_xref="GI:9758814"

CDS

/translation="MKDVLIVYSLIKMTKLRSCLDCCNSSTYTLQKHLKNEHDG
 LVLLWVQSQMRCCEMKRHYEVVDLMSATEPHIAEVVHHQLPYLDTLILFKLIS"
 join(18792..18841,19220..19412,19494..19579,19875..19953,
 20464..20507,20697..20783,20894..21006,21097..21152,
 21216..21308,21350..21509)
 /note="gene_id:MX110.5
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09349.1"
 /db_xref="GI:9758815"

CDS

/translation="MNSGPKQISDORGSHRHHPFQIIHLIGNFFRIKWSYKIRVLN
 QTGAPVULFECCLVPSSIFLIIOTKPKGRALSNQOIVPSLINGYIALYFLW3KG
 LKSCGPLRALISLSEYGAIVGLSGVLRGRGHVMMKLPTDSTETKEEFVQTEAL3G
 M3MIPVAFGILSALRRVIARVSLKQOKRLHAITITATCFGLFFVAMMDIIGSS
 SCKTSELPEFAWFLSTIIIGIILIFVYDNIABERLHMVFPSPRLMWACVCIIVDI
 RVANIK3NISSTQEERLEFVIRSRISKRNPCSPFIPPYINDVDVLNVLDQVNL3GFI
 R3IPVPSFF3FCFLGEAYANILV"
 join(26073..30268,30416..31552)
 /note="gene_id:MX110.6"
 /pseudo

CDS

/codon_start=1
 /evidence=not_experimental
 /product="retroelement pol polyprotein-like"
 3202..33620
 /note="contains similarity to retroelement pol
 polyprotein
 gene_id:MX110.7"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09350.1"
 /db_xref="GI:9758816"

/translation="WVEGTRINKRSKTMSTPQSEPAVHTWKEDKKALKVLSPLP
 SEPTFAREILEKMGLELIDEFTKNGLGAFSDMDHTFSEPTVDFSTMEYTTIN
 PPIPLAKGCI...NFKVTKAFSISIDPCEAYGFENKSEMFPPKFGIHFVWELASGK
 F3CNASIKSRHPVIRYALKLHALFQGETSTTVSECMFPGQKELLEDADEN
 GGRVLHDOGLNDVNYGICFASYLEDKRNLIKGTKSIVGGTLTFLFKAGVDLSP
 ENALPKREYIDYENTVRSCSLKSDDAFIYLTIDRENDKLECLILPSKEILNVE3EDI
 MFLPSKYOVPNLQPMIEDEDLPPSPRODPYNLKNCDLPSTFVDPKTKMEXLOS
 IK3QRKINKKSYALGKLLKKVKLQPPDYVSSDELDLVFFSNEEDRDNTLES3ENNE
 TPAISSSRKRNAROH3YMP"
 join(36310..37248,37409..37549,37633..37824,38363..38519)
 /note="embicAB86928.1
 gene_id:MX110.8
 similar to unknown protein"

CDS

/pseudo
 /codon_start=1

/evidence=not_experimental
 /join(39564..40460,40627..40767,40853..41185)
 /note="db:AAD55600.1
 gene_id:MX110.9
 strong similarity to unknown protein"

CDS

/codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09351.1"
 /db_xref="GI:9758817"

/translation="MDLLSNLPYELLCILHSFLTKEAALT3VLSKRWRLIAFVPRV
 DIVDHDIRELDFDVRVLAQNSPINKFSLDSCGVSDSDRVDWCIONWVRGVSELN
 LISFVDSVFEDDYNLFPVFNKKNLKLGSYISWLGSIPLMKLTVLSVLSVSE
 KFEILLHALPALLEEIVMNNIYKPRD3VVDLIVSSESLKTLTINFIYCTHPLSF3DIPIL
 AYLSVSGYVDDYAEAKMENLPEARISLIVYEGDIS3R3RALINNDLEDEYDLOPE
 NVKLMJGIRNIRCLYLSPTNLEVLCCESHPVFNKLSL3K3SLK3SADWQ3APVLL
 RNCPHLETLYL3GLLHHVTDKGDAC3D3REDGRSLT3CPKVLEIKGFG3T3KEM
 NMIKHFLE3YF3CL3KEL3MYMEEN3PAQ3UR3PEVIT3EM3RYK3K3SS3CN3VQ3L3V3G3KL
 YKKWTP"

```
CDS      join(42837..43766,43859..43975,44083..44493)
        /note="emb|AR86928.1"
        gene_id:MX110.10
        similar to unknown protein"
        /pseudo
        /codon_start=1
        /evidence=not_experimental
CDS      join(45089..45677,45843..47158,47312..47452,47544..47871)
        /note="contains similarity to non-UTR retroelement reverse transcriptase"
        gene_id:MX110.11"
        /pseudo
        /codon_start=1
        /evidence=not_experimental
CDS      48978..49433
        /note="gene_id:MX110.12
        unknown protein:"
        /codon_start=1
        /evidence=not_experimental
        /protein_id="BAB09352.1"
        /db_xref="GI:9758818"
        /translation="MHVDSMKRARSVAVEGKERSSNQDETREOSKHAKTAKAR
        MNVXASMPLEPKKKSLMILSMKXEDTODQNNVTVVEKESKEKQVQNTRE
        NAEMLSKVEVWPHWLYEEBWTWPAFGWGWQRYDAVMGSKSFN"
        complement(join(50633..50972,51043..51177,51269..51438))
        /note="gene_id:MX110.13
        sp|P10798"
        /codon_start=1
        /evidence=not_experimental

Query Match      31.8%; Score 540.2; DB 8; Length 83646;
Best Local Similarity 93.6%; Pred. No. 26-84;
Matches 574; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

y  i  TGGGTTTTATGATTAACATGACAAATATTTATTATTCATGAGCTTTTATTGAGAG 60
b  23771 TGGGTTTTATGATTAACATGACAAATATTTATTATTCATGAGCTTTTATTGAGAG 23712
y  61  CATCAGAAATATTAATATCATGCTTTAATACATCTTTTGTCTTAAATACATGCAAT 120
p  23711 CATCAGAAATATTAATATCATGCTTTAATACATCTTTTGTCTTAAATACATGCAAT 23652
y  121 TTAATAATCAGACATCTTGTCTTAAATCAATCTTAATCTTATATACAGCAATTCAT 180
b  23651 TTAATAATCAGACATCTTGTCTTAAATCAATCTTAATCTTATATACAGCAATTCAT 23592
y  181 GSAAATTCAGGTAAAGAGAAATTAAGAATGAGAGATGAGAGATTCATGCAAA 240
b  23591 GSAAATTCAGGTAAAGAGAAATTAAGAATGAGAGATGAGAGATTCATGCAAA 23533
y  241 AGAAGAGAGAACATGATGATGAGCAACAAATTAAGAGATATGATGATATTTATCAAG 300
b  23532 AGAAGAGAGAACATGATGATGAGCAACAAATTAAGAGATATGATGATATTTATCAAG 23473
y  301 GTGGTGAAGATTTATTAGAGAGGGAGAGAGAAATAGAAAAGAGAAATGACATGCTGA 360
b  23472 GTGGTGAAGATTTATTAGAGAGGGAGAGAGAAATAGAAAAGAGAAATGACATGCTGA 23413
y  361 TCTGAAGAAGATGAATTTGTCTTAAAGATGAAGAGAGAGAACTCCATGCTAAAGTC 420
b  23412 TCTGAAGAAGATGAATTTGTCTTAAAGATGAAGAGAGAGAACTCCATGCTAAAGTC 23354
y  421 TCGTAAGCAATGTAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
b  23352 TCGTAAGCAATGTAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23293
y  481 TTAATTTGCAAAATTTCTGTAGCCGACAAATTAATTTGTGCTCAAGGTTATTTCG 540
b  23292 TTAATTTGCAAAATTTCTGTAGCCGACCAATTAATTTGTGCTCAAGGTTATTTCG 23233
y  541 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATATCTTAATAAATATGCGGATACCAA 600
b  23232 TATCTTTTGAAGTCAAAAGTTATTTCTTACATATATCTTACATATGTTTGTGTTAAAT 23173
```

QY 601 TTCTTCCACACAT 613
I
DL 23172 TTATTTAAATAT 23160

RESULT 4
AC012394
LOCUS

DEFINITION
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence,
complete sequence.

ACCESSION
AC012394
VERSION
AC012394.3 GI:6554469
KEYWORDS
HTG.
SOURCE
Arabidopsis thaliana.
ORGANISM
Arabidopsis thaliana.

Biological Process: Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
AUTHORS

Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Renning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence
Unpublished

REFERENCE
AUTHORS

Lin, X. and Kaul, S.
Direct Submission
Submitted (27-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org

REFERENCE
AUTHORS

Lin, X.
Direct Submission
Submitted (10-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

REFERENCE
AUTHORS

Town, C.D. and Kaul, S.
Direct Submission
Submitted (12-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

REFERENCE
AUTHORS

Town, C.D. and Kaul, S.
Direct Submission
Submitted (05-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

REFERENCE
AUTHORS

Town, C.D. and Kaul, S.
Direct Submission
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org

COMMENT

BAC clone F15M4 is from Arabidopsis thaliana chromosome 1
the orientation of the sequence is from SP6 to T7 end of the BAC
clone.

Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of Glimmer, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm), and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR

{http://www.tigr.org/tdb/tgi.shtml}. Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by

tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
Simple repeats are identified by repeatmasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

```

source
1..9487
  /organism="Arabidopsis thaliana"
  /cultivar="Columbia"
  /db_xref="taxon:3702"
  /chromosome="1"
  /clone="F15M4"
  complement(967..1017)
  /rpt_family="AT-rich"
  complement(1289..1336)
  /rpt_family="AT-rich"
  complement(14053..14080)
  /rpt_family="AT-rich"
  complement(16943..19510)
  /gene="F15M4.1"
  /note="identical to hydroxy methylglutaryl CoA reductase
  (AA 1-592) GB:X15032 [Arabidopsis thaliana]"
  complement(join(<6943..7515,7507..7953,8115..8296,
  8522..>9510))
  /gene="F15M4.1"
  complement(join(7255..7515,7507..7953,8115..8296,
  8522..9510))
  /gene="F15M4.1"
  /codon_start=1
  /product="hydroxy methylglutaryl CoA reductase (AA 1-592);
  9510-7255"
  /protein_id="AAF16652.1"
  /db_xref="GI:6554470"
  /translation="MDLRPRPKPPVITNNNSGFSYQPTSDCRRERNTIAP
  FKASALPLPLYNVFFLFFSVAYLLHWRDKIRYWPVHVYIIELGAILALI
  APVILLGFGFQVQSIRASQAKLAUTIDDDHRLVTCSPPTFVSVAKLPP
  EPIVLLPEDEETVSDIVGVIPSVLSLEKGCXRAASIRREALQRTVGSIEGL
  PLDGDYSELGQCCMPVGIQIPVGIAGPLLDGYEYSVPMTTECCIVASTNRGC
  KAMFLSGGATSTV.KDGMTAPVVRFSARRASBLKFLFENPFDILAVVFNESSPE
  ARDSQKPAVNIKAGKAYVFCSTGDAMGMVSKGVNLEVLTDFDPMVDVIGISGN
  PCSQKPAVNIKAGKAYVFCSTGDAMGMVSKGVNLEVLTDFDPMVDVIGISGN
  SLGGNNANSTVSVFATGQDPQNVSSOCITMNEANGDKDHI:SVYMSIEVQ
  TVGGQIQAASACNLGLGVGASTESPGMNRRLATIVAGAVLAGELSLMSATAAQ
  LVRSHMKYNSRQISGATTTTTTT"
  complement(7258..7283)
  /rpt_family="(CAA)n"
  complement(11154..11209)
  /rpt_family="AT-rich"
  complement(11551..11580)
  /rpt_family="AT-rich"
  complement(11933..12004)
  /rpt_family="AT-rich"
  12390..13268
  /gene="F15M4.2"
  /note="similar to unknown protein GB:AAD30653 [Arabidopsis
  thaliana]"
  join(<12390..12391,12491..12629,12730..12861,12929..13019,
  13099..>13268)
  /gene="F15M4.2"
  join(12390..12391,12491..12629,12730..12861,12929..13019,
  13099..13268)
  /gene="F15M4.2"
  /codon_start=1
  /product="unknown protein; 12390-13268"
  /protein_id="AAF16653.1"
  /db_xref="GI:6554471"
  /translation="MFQKALYVKGELWRETIPLCSGSKYKLGKLSKNSWYEVKISY
  PASIPARVSIQ:LKNEVEKLQNMRLNTEK:LFAKFS:KEVNNKAG:KNVILVLEP
  EGVALPNSRFRVFIIVYCEQLMGIPYSSWSVLLVALCLVSVFVPSLPSLLI
  TRQQLNSHRSHGKDS"
  14056..15506
  /gene="F15M4.3"
  /note="similar to cinnamoyl-CoA reductase GB:CAA56103
  [Eucalyptus gunnii]"
  join(<14056..14170,14277..14431,14516..14701,14819..14990,

```

```

15096..15285,15362..>15506)
/gene="F15M4.3"
join(14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..15506)
/gene="F15M4.3"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 14056-15506"
/protein_id="AAF16654.1"
/db_xref="GI:6554472"
/translation="MAYKQKVCVTGAGGFIASMLVKFLLSRGVIYVHGIVRUCDEKND
HLKLDNAKNLKLKADLPDEGLFSNIDGCSGVFIASVPVPEGELIKPALIGTK
NVLACTETKVKVYVYSSIAAYVNFNPQDVAKEDCSQYDLSEGSJWHYIFL
AKTLTERALEMSKRNFAVYVTLCPVSTIPRLOSILNSSLGLLRFKIGKISLLSD
EYLYVDVRDADALLVYENREATYICNHSLSYIDLSMEKLNMYPKRNFPESTE
VKEKEVRPLSAEKLKLNKGKFLPELTIDTSVWSFEAAGDLPKA"
16281..16361
/rpt_family="(CAI)r"
15955..18423
/gene="F15M4.4"
/note="Contains Pfam profile: PF00076 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
join(<16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..>18423)
/gene="F15M4.4"
join(<16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..18423)
/gene="F15M4.4"
/codon_start=1
/product="putative RNA-binding protein; 16955-18423"
/protein_id="AAF16655.1"
/db_xref="GI:6554473"
/translation="MAYQVPVSGFHYLNSPFGDTTKVYVGLAMPTQSETLRQHF
EYGEILCAVVIADKNTGSKGYGFVFRDEPAARRACADPTPIIDRRANCKLASLG
RRPPLPFAVNPMDLLPRILEMCRVLEHLEIRISNHFHIIISKELSLCHVTAY
GREYVNSQGLYSPYMQQYLQVYGVGAVNSPVYQYQLSOTIPNGHGYTAVQVYS
VGRSHILQLGPTVSTMTTSSMPALQAPYPSGIPGAPVQSHIIVHSPQMGSTASQ
"TR"
complement(19352..20843)
/gene="F15M4.5"
complement(join(<19352..19378,19455..19493,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))
/gene="F15M4.5"
complement(join(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))
/gene="F15M4.5"
/codon_start=1
/product="unknown protein; 20843-19352"
/protein_id="AAF16656.1"
/db_xref="GI:6554474"
/translation="MAAISPWLSQSFNSPRVITIDSRCCSSISAISVLDSNNEQ
HRISSRDVGMKRRQVLMQIASSVFFLPALISPAFAINASEAFRVYTDENKFEISL
POVSYATIGLGPDPFTMESFGKVEPAFTLYSGLDRSQKQPVGVYAKLIDRSKRGY
FIYETLQNPGEARKHLYSAIGMATNGWYNRLLYTVTGFTDEESAQSISKQIKYKSPR
FI"
complement(20998..21026)
/rpt_family="AT-rich"
complement(21140..21167)
/rpt_family="AT-rich"
22462..23090
/gene="F15M4.6"

```

```

Query Match      31.4%; Score 534.6; DB 8; Length 94487;
Best Local Similarity 96.5%; Pred. No. 1.8e-83;
Matches 557; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

Cy 1 TGGGTTTATTGGAACACATGACAAATATTTATTATTTCATGAGTTTTTTATTCGATAG 60
    |||||
Db 1917 TGGGTTTATTGGAACATGACAAATATTTATTATTTCATGAGTTTTTTATTCGATAG 1976
    |||||

Gy 61 CATGACAAATATTAAATATCATCAGTGTTAATACATGTTTGTCTTAATACATCGCAT 120
    |||||

```

```
20 1977 CATGACAAATATTAATATATTAATGTTAAACAACTGTTTCTTCTTAAATACATGCAAT 2036
21 121 TTAATAATCAGACATTTGTTTAAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 180
22 2037 TTAATAATCAGACATTTGTTTAAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2096
23 181 GGAATATTCAGGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
24 2097 GGAGATTCAGGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2156
25 241 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
26 2157 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2216
27 300 GTTGCTGAGAGATTTTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
28 2217 GTTGCTGAGAGATTTTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2276
29 360 ATCTGAAGAGAGATTTGTTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
30 2277 ATCTGAAGAGAGATTTGTTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2336
31 420 CIGCTAAAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
32 2337 CIGCTAAAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2396
33 480 ACTACTATATGCAAAATTTCTGTAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
34 2397 ACTAAATATTTGCCAAATTTCTGTAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2456
35 540 GTATTCCTTTGAGTCAAAAGTATTTCTTCTTCAATATA 576
36 2457 GATTCCTTTTCAACACAAAGTATTTCTTCTTCAATATA 2493

RESULT 5
C015450/C
OCUS
Arabidopsis thaliana chromosome 1 BAC F14G5 genomic sequence,
complete sequence.
AC015450
AC015450.5 G:12323969
HIG.
Arabidopsis thaliana.
Arabidopsis thaliana
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: Eudicotyledons: Core eudicots:
Rosidae: eurosids I: Brassicales: Brassicaceae: Arabidopses.
1 (bases 1 to 100806)
Lin.X., Kaul.S., Town.C.D., Benito.M.-I., Greasy.T.H., Haas.B.J.,
Wu.D., Maiti.R., Roeding.C.M., Koo.H., Fujii.C.Y., Utterback.T.R.,
Barstead.M.E., Bowman.C.L., White.O., Nierman.W.C. and Fraser.C.M.
Arabidopsis thaliana chromosome 1 BAC F14G5 genomic sequence
Unpublished
2 (bases 1 to 100806)
Lin.X. and Kaul.S.
Direct Submission
Submitted (16-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 100806)
Town.C.D. and Kaul.S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280761.
Address all correspondence to:at@tigr.org

BAC clone F14G5 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mt.edu/GENSCAN.html), GenemarkEM (Mark Borodovsky,
```

```
http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmer_hgm/glimmerm.htm), and
Geneslicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are annotated by
tRNAscan-SF (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SF/).
Simple repeats are identified by RepeatMasker (Arian Smith,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Location/Qualifiers
1..100806
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F14G5"
misc_feature
1
/note="35000 nt before this point were not included in
the submitted sequence, due to overlap with another BAC
(F15M4)."
5217..5774
/gene="F14G6.1"
/note="contains Pfam profile: PF00097 zinc finger, C3HC4
type (RING finger)"
5217..5774
/gene="F14G6.1"
5217..5774
/gene="F14G6.1"
/codon_start=1
/product="putative RING zinc finger protein; 5217-5774"
/db_xref="GI:12323975"
/translation="MARLLFLLQEA NSTSPAEASPPFNSDLVLAVLCAITCIG
TAVSRCAWLRIASRNQDOTHPVPAANKGLKVKVLSRLPKLTYSPOSPFAKLV
PCAICCTEFAAGDELRLVLPQCCHGHCVCIDTILGSHSCPCSCGRIJVIWIRCHQKGL
PGSSSSGPEPDTRIKQREDGPNLP"
complement(8331..10137)
/gene="F14G6.2"
/note="N-term similar to N-term of NAM GB:CAA63101
:Petunia x hybrida [apical meristem formation], CUC2
GB:BAA19529 [Arabidopsis thaliana], GRAB2 protein
GB:CAA09372 [Triticum sp.]"
complement(join(<8331..8855,9526..9800,9933..>10137))
/gene="F14G6.3"
complement(join(8331..8855,9526..9800,9933..10137))
/gene="F14G6.2"
/codon_start=1
/product="unknown protein; 10137-8331"
/db_xref="GI:12323982"
/translation="MMLAVEDVLSELAGEERNEGLPPGFPHUPTDEELTFYLASKI
PHGLSGIHISEVDLNKCEPEHPELMKGRERWYFSLDRKPYGLRNRATYAG
WKATQKREVFSGGGQGLVGMKLTLYFKGRAPGLTKWMEYLRLENDHSHRTCK
FEWTCRVFNKTDGRKNVGLIHQISYLNHNSLTHHHEALPLLIEFNKILTNF
PSLYDDPHQNNNNFLHSSGHNIDELKALINPVVQLGIFPSGNNNDGDF
FNLGKVTQSSNGNEIDVRYLENPLFQEA SYGLGSSSPGRLMLDSCPGLGQL"
complement(15417..18176)
/gene="F14G6.3"
/note="similar to phosphate transporter GB:BAA20522
[Catharanthus roseus]; contains Pfam profile: PF00083
Sugar (and other) transporter"
complement(join(<15417..16305,17467..>18176))
/gene="F14G6.3"
complement(join(15417..16305,17467..18176))
```


		MKV...EYVAMNGLPFCPCGGH.LDIISEKDGAKRYECTIKYNKNJLHIRAKDKK
exon		MVEFNRRLREOVNDNHQIQISLEYNQMRSKIDELCKKHGTIWRICCGE"
		/complement(28303). .28494)
		/gene="A74g08720"
		/number=1
intron		/complement(28495). .28564)
		/gene="A74g08720"
		/number=1
exon		/complement(28565). .28761)
		/gene="A74g08720"
		/number=2
intron		/complement(28762). .29625)
		/gene="A74g08720"
		/number=2
exon		/complement(29626). .30162)
		/gene="A74g08720"
		/number=3
intron		/complement(30163). .30332)
		/gene="A74g08720"
		/number=3
exon		/complement(30333). .31953)
		/gene="A74g08720"
		/number=4
gene		34618. .36007
		/gene="A74g08730"
gene		/complement(join(34618. .34852),35539). .35690,35628). .36007;
		/gene="A74g08730"
CDS		/complement(join(34618. .34852),35539). .35690,35628). .36007;
		/gene="A74g08730"
		/codon_start=1
		/product="hypothetical protein"
		/protein_id="CAB82109.1"
		/db_xref="GI:7321062"
		/translation="NEGTFNFQVMETKVDLLAKVQD.LTSKMQLCATNYDLLTSSKNV
		.LCATNSDLKANMITSKKKXVPAAKKTILANAQSDSESSDDIELCPRIKLIVIG
		LRSYFIILLFAIDLEKGNAKALEINDGDMAGKELVIALINRDTSLGTSCNGSGLOG
		RFGGRFGVGGIPIGGRGOGGLGYTYAQV"
exon		/complement(34618. .34852)
		/gene="A74g08730"
		/number=1
intron		/complement(34853). .35536)
		/gene="A74g08730"
		/number=1
exon		/complement(35539). .35690)
		/gene="A74g08730"
		/number=2
intron		/complement(35691). .35827)
	Query Match	29.4%; Score 499.6; DB 8; Length 104386;
	Best Local Similarity	90.68; Pred. NC. 1.9e-77;
	Matches 557; Conservative	0; Mismatches 52; Indels 6; Gaps 27;
QY	1	TGGGTTTTATTGGATAACATGCACAATAATTTATTTATTTTCATGAGTTTATTTCGGATAG 60
Ddb	87736	TCGGTTTATTGGATAACATGACTAATATTATATATTTCAIGAGTTTCTTGTSNAG 87677
QY	61	CATGCAAAATTAATAATACAGTGGTAATAACATGTTTGTGTCTTAATAATCATGCAT 120
Ddb	87676	CATGCAAAATTAATAATACAGTGGTAATAACATGTTTGTGTCTTAATAATCATGCAT 87617
QY	121	TTAAAATCAGACATTGTGTTTAAAAATCAAATCTAATCTCTATATCACACACACATTGAC 180
Ddb	87616	TTAAAATCAGACATTGTGTTTAAACATCAAAATCTAATCTCTATATCACACACACATTGAT 87557
QY	181	GGAATTCAGGTAAAAGAGAAAATAAAGAATGAGAGATAGAGAGATATTCATGSAAA 240
Ddb	87556	GGAGAAATTCAGCAAAAAGAGAAAACAAGAATGAGAGATAGAGAGATTTCTATGSAAA 67497
QY	241	ACAAGACAGACACATGTAGGTGACAAAAATAAGACAGATATGATGATATATTTATGAGAS 300
Ddb	87496	ATAAGAGGGGAACATGTGGGTGACAAAAATAAGACAGATATGATGATATATTTATGAGAG 67437
QY	301	GTGGTGAAGATTTATTT-----TAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATG 355

DB	B7436		CTGGTCAAGATAATCTTAGGAGGGGAGAGAATAGAATAGAAAAGAAAAACGCATG	87377
QY	356		GTCGAATCTCGAAGAAGTAATGTGGTTAAAGATGCAAGACAGAAAAGAGAAGACTCCATGGCTA	415
DB	B7375		GTCGAATCTCGAAGAAGTAATGTGGTTAAAGATGCAAGACAGAAAAGAGAAGACTCCATGGCTA	87317
QY	416		RAGTCTGTGAAGAAGATGAAAAGAAAACAAGAAGAGGAGAGAGAGAAGAGAGGCTAAA	475
DB	B7316		RAGTCTGTGAAGAAGATGAAAAGAAAACAAGAAGAGGAGAGAGAGAGAGGCTAAA	87257
QY	476		ATAGACTAACATTGCCAAAATTCCTGTAGCACCAGAACTACTATTGGTCCAAGGTTAT	535
DB	B7256		ATAGACTAACATTGCCAAAATTCCTGTAGCACCAGAACTACTATTGGTCCAAGGTTAT	87198
QY	536		TGCTATTCITTTGAAGTCAAAGTATTTCITTACATATACCTCTAAAATAATAGCCCAT	595
DB	B7197		TGCTATTCITTTGAAGTCAAAGTATTTCITTACATATACCTCTTTTTATATAAAAAAAT	87138
QY	596		AACAATTTTCCACA	610
DB	B7137		ATAACGTTTGCAA	87123

RESULT 7
 ATCHRV25/C
 LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 25.
 ACCESSION AL615813
 VERSION AJ1615.3.2 GI:7267524
 KEYWORDS Arabidopsis thaliana.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.

REFERENCE 1 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Brassicales; Brassicaceae; Arabidopsi-
 s [bases 1 to 26363; 134758 to 179771]
 Spegel,L.A., Huang,E.N., Nascimento,L.U., de la Bastide,M.,
 Vil,D.M., Preston,R.R., Matero,A., Shah,R., O'Shaughnessy,A.,
 Rodriguez,M., Shekher,M., Schutz,K., See,L.H., Swaby,I.,
 Habermann,K., Dedhia,N.N., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
 Unpublished
 2 [bases 25642 to 34296]
 Robben,J.J., Grymonprez,B., Bastiaens,I., Voickaert,G., Mewes,H.W.,
 Lemcke,K. and Mayer,K.F.X.
 Unpublished
 3 [bases 33706 to 33707]
 Lamar,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and
 Mayer,K.F.X.
 Unpublished
 4 [bases 1 to 179771]
 EU Arabidopsis sequencing project.
 Direct Submission
 Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 19a, D-82152 Martinsried, FRG. E-mail:
 lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bsrc.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
 this fragment has an overlap with AtCHRIV26 at the 5' end and an
 overlap with AtCHRIV25 at the 3' end.

FEATURES
 source
 location/Qualifiers
 1..179771
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 8902..10347
 gene
 CDS
 join(8902..9093,9193..9553,9686..10347)

```

/gene="AT4g08620"
/notes="similarity to various hypothetical proteins of
Arabidopsis thaliana Chromosome II, III and IV"
/codon_start=1
/product="putative protein"
/protein_id="CA878007.1"
/db_xref="GI:7267525"
/translation="MSVSNKSGVGGTGLKFNVIDAFVRERVRLEFPNCEKPPVIT
IGEEVLEAMNGIAKRCMIUKFFPIRELEBEYLAAITGGIRKRVFGKYLIVGCLNSPFD
PLDDAVTTPVWYRLNIPNYPYHRLLEBEIAGLKLKLVDLNIITFGKGFARVCZ
EVNLAKPLGTVLINDRIYFAVEEVEEGFTVVGSGIRVERPAQRKAVFAASGGRS
NRSLEPLKNGQVEIENRFGLEEDMVSAPIGEVALEGSNKENKYGKXNMKRESGV
QVKEACQVNTFKKGCGKNGKPKWRNGGLIMEPNGPRCKQGSVKNKPVRGITFGPIK
GESSNLVGDAISVAGKRLRQVRSGRPGGAFVSGLETRSGEMKMSMELPQNNHQQC
SKELLCDTAIVV"
exon: 5932..9093
/gene="AT4g08820"
/number=1
exon: 9094..9192
/gene="AT4g08820"
/number=1
exon: 9193..9553
/gene="AT4g08820"
/number=2
exon: 9554..9685
/gene="AT4g08820"
/number=2
exon: 9686..10347
/gene="AT4g08820"
/number=3
gene: 10891..11438
/gene="AT4g08830"
CDS: join(10891..11025,11244..11393,11430..12120,12151..13089,
13162..13585,13634..14348)
/notes="similarity to various putative reverse
transcriptases of Arabidopsis thaliana
Contains Anticacy-Transfer RNA synthetases Class-II
signatures AA699-708"
/codon_start=1
/product="putative protein"
/protein_id="CA878008.1"
/db_xref="GI:7267526"
/translation="MSPDSLEFGIEWINRVSLDMGTSGNKYTKRGRVENTFAKRLD
RLAKNRDVFQDKKKEEMREIOVVOADLVQSDMDLKRKEEDLIKAFDKSREKWI
ALGDRNTNYPTITILRRPNRVERLHNEEGTWIFDAKLEALAVQYQSLYSMLCVL
QVYNPLPDGSGERLRETVSLNKPTELATEVEVAIRSMGKYKAPGPDGQPIFYQGW
EVGNSVIRPALDFFSSGILPPQTDNALVVJIPKVPKPERMQFRPISLCAIIFKMIT
KMVLRKKVIGKLGISQSPKISHGFADDLILFAEASVQIRVIRITIEFICIASQKVS
AVAKKKSIGLSQGGPKISHTFADLLFAEASVQIRVIRITIEFICIASQKVS
LDKSKIPFSKNSRDLEKLTKSGSGIKSTRELKYLGMPLORRINKDTFGYDERVS
SRLAGNKGRLSPAGRLTLKSVLSLPIHTMSTISLPQSTLEGLDKLAPVPLKSSA
EKKLHJVANDRVCPLPKSEGGILRTSKCNKALVSKVGRHLINDRYSLWRLRSKY
RVGLREVSGSRKSVNGSDIIPSDNWLSEALINRAVEIPNSKELRVRKDEKAN
GIWGLKDKIPY-SYHTRLEAAVVDVSGARDLSWGSVADGVTFSKAYKLTED
HDPRPNNAAEFDRIWRVALLERYKFIWHITGDSVQCKGQDETILHVKCPQSAG
IMRRIQVQVSQYDFNCSIPGMLVYNLGMKNATGYAWTLFAIVWMSKMKSCGYVE
GEVKKCRDRYKFKRLAAEVSXHAHATHSNGGLRTRVERLVANKPPDGEWVXINIDGA
SGNLGLATTGGVLRDGLGHWCQGFALDGVCSNPLAEJMGVYVGLYMAHERRTIVE
LEVDSELVGLTGTGSDTHSLSLFVRLKHGFL"
exon: 10891..11025
/gene="AT4g08830"
/number=1
intron: 11026..11243
/gene="AT4g08830"
/number=1
exon: 11244..11393
/gene="AT4g08830"
/number=2
intron: 11394..11425
/gene="AT4g08830"
/number=2
exon: 11430..12120

```

```

/gene="AT4g08830"
/number=3
intron: 12121..12360
/gene="AT4g08830"
/number=3
exon: 12361..13089
/gene="AT4g08830"
/number=4
intron: 13090..13161
/gene="AT4g08830"
/number=4
exon: 13162..13585
/gene="AT4g08830"
/number=5
intron: 13586..13633
/gene="AT4g08830"
/number=5
exon: 13634..14348
/gene="AT4g08830"
/number=6
repeat_region: 16002..16053
gene: 18436..120750
/gene="AT4g08840"
CDS: join(18436..19424,19870..20032,20124..20287,20492..20750)
/notes="similarity to probable regulatory protein Y3L014w
Saccharomyces cerevisiae, PIR2:S64016
Contains Prokaryotic membrane lipoprotein lipid attachment
site AA82-92"
/codon_start=1
/product="putative protein"
/protein_id="CA878009.1"
/db_xref="GI:7367527"
/translation="MDGFFPGDLRQGSFTDGLGNGPPLTSSVNGHFHSGRTINP
FLNLRKLDTSLMADGVDMGLCONLSKMSIDERSNFNHSFSYGYQGRESFHG
EASSGMRFVGVGDYHREPDQLRVASFGESAMSYVDGSDYHRLPFLALQFASNP
NPRCTENMSLLNRDYMLEHFNQOIRRDFSLVPKSP-LAFHEERILPPFSAMGCSR
ELDGADCKMKNKEDSLDLASVYVLMADGLGRLLQKLFYDEGNFVDMVILFK
EVNNVIELGTDPEGNLIQKLIENCNEFORTQILRLTSKPLAVKLSINNYGIRVV
QKLIETVTKQISLVKVSALVFGFLSLFRELNGNHVILNCLFFSPDNKFLBAEK
FCIEATIRHGCCVLRQVSYSVGEQHEKLVDEISNLSLLAODPFGNLTGVYIIIEKK
VGVNVVLFELRGNVYKLTATQKFGSHVVEKLRYPESRSQIVNE-VSVLNFGLYLODP
YANYIQCLASKTIVS"
exon: 18436..19424
/gene="AT4g08840"
/number=1
intron: 19425..19669
/gene="AT4g08840"
/number=1
exon: 19870..20032
/gene="AT4g08840"
/number=2
intron: 20033..20123
/gene="AT4g08840"
/number=2
exon: 20124..20287
/gene="AT4g08840"
/number=3
intron: 20288..20491
/gene="AT4g08840"
/number=3
exon: 20492..20750
/gene="AT4g08840"
/number=4
exon: 23831..27580
/gene="AT4g08850"
complement(join(23831..24120,24787..27580))
/gene="AT4g08850"
complement(join(23831..24120,24787..27580))
/notes="AT4g08850"
/notes="Est GB:F14113 is reverse and marks the end of the
gene
similarity to receptor protein kinase-like protein -

```

Arabidopsis thaliana, P12.T05098
Contains Prokaryotic membrane lipid attachment site AA:2-22:Protein kinases signatures and profile AA763-784:Protein kinases signatures and profile AA883-895
Contains EST ab:F14412, Z17737, R84191, T22574, AA394763, AA395692, F1413*

Query Match 29.4% Score 499.8; DB 8; Length 179771;
Best Local Similarity 90.6%; Pred. No. 1.7e-77;
Matches 557; Conservative 0; Mismatches 52; Indels 6; Gaps 2;

QY 1 TGGGTTTATTCGATACATGACAAATATTATTCATCATGATTTTATTCGATAG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17646 TCGGTTTATTCGATACATGACAAATATTATTCATCATGATTTTATTCGATAG 17597
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CATGCAAAATATTAATATATATATCATGTTTAAATACATGTTTGTCTTAAATACATGANT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17586 CATGCAAAATATTAATATATATATCATGTTTAAATACATGTTTGTCTTAAATACATGANT 17527
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 TTAATATCAGACATTTTGTCTTAAATACATGTTTGTCTTAAATACATGANT 160
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17526 TTAATATCAGACATTTTGTCTTAAATACATGTTTGTCTTAAATACATGANT 17457
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GGAATATTCAGTTAAAGAGAAATTAAGATGAGATGAGATGATTCATCGAAA 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17466 GGAGATTCAGTTAAAGAGAAATTAAGATGAGATGAGATGATTCATCGAAA 17427
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 AGAAAGAGACATGATGATGACAAATAAAGAGATGATGATGATGATGATGATGATGATGAT 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17406 ATAAAGAGGAGACATGATGATGACAAATAAAGAGATGATGATGATGATGATGATGATGAT 17347
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GGGTCAAGATTTT-----TAGGAGGGGAGAGAGAAATAGAAAGAAATACATG 355
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17346 GGGTCAAGATTTT-----TAGGAGGGGAGAGAGAAATAGAAAGAAATACATG 17287
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 356 GTGAATCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 415
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17286 GTGAATCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 17227
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 416 AGTCTCTGTAAGACATGAAAGAGAAATAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17226 AGTCTCTGTAAGACATGAAAGAGAAATAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 17167
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 476 ATAGACTAATTCGCAAAATTCGTATGACCAAAATAGATGATGATGATGATGATGATGATGATGAT 535
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17166 ATAGACTAATTCGCAAAATTCGTATGACCAAAATAGATGATGATGATGATGATGATGATGATGAT 17106
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 536 TTGCTGATTCCTTTGAGTCAAAAGTATTTCTTACATATCTTAAATATACCCAT 595
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17107 TTGCTGATTCCTTTGAGTCAAAAGTATTTCTTACATATCTTAAATATACCCAT 17045
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 596 ACCAATTTTCCACA 610
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17047 ATAACGTTTGCAAA 17033
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS AF071527 95579 bp DNA linear P12 06-MAY-1999
DEFINITION Arabidopsis thaliana BAC F9H3, from chromosome IV near 18.8 cm,
complete sequence.
ACCESSION AF071527
VERSION AF071527.1 GI:3236479
KEYWORDS HTG.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 95519)
Huang,E.N., Parnell,L.D., de la Bastide,M., Schutz,K.,
Habermann,K., Dedhia,N.N. and McCombie,W.R.

Genomic sequence of Arabidopsis thaliana BAC F9H3, chromosome IV,
18.8 cm
Unpublished
2 (bases 1 to 95519)
Huang,E.N., Parnell,L.D., de la Bastide,M., Schutz,K.,
Habermann,K., Dedhia,N.N. and McCombie,W.R.
Direct Submission
Submitted (10-JUN-1998) Cold Spring Harbor Laboratories, Lita
Annenberg Hazen Genome Center, 1, Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 95519)
Parnell,L.D. and McCombie,W.R.
Direct Submission
Submitted (31-DEC-1998) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
Arabidopsis thaliana BAC F9H3 from chromosome 4, near 18.8 cm
4 (bases 1 to 95519)
Parnell,L.D. and McCombie,W.R.
Direct Submission
Submitted (01-FEB-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.5 cm
5 (bases 1 to 95519)
Parnell,L.D.
Direct Submission
Submitted (06-MAY-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724
Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.8 cm
BAC F9H3 is assigned between YACs C14A7 and C1C8B1 and maps to
near 18.8 cm on the Lister & Dean RI map. Position 1 of F9H3 is
oriented toward the centromere and position 95519 is oriented
toward the telomere. For more information on the mapping,
sequencing and annotation of F9H3, please see
<http://www.cshl.org/arabweb/F9H3-titlepage.html>. A graphic view of
our annotation is also available at this url. Gene models are built
with exons predicted by GENSCAN
(<http://CCR-061.mit.edu/GENSCAN.html>), MZEF
(<http://www.cshl.org/genefinder>) and GRAY,
(<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites
predicted by NetPlantGene
(<http://www.cbs.dtu.dk/netgene/cbsnetpgene.html>). Alternate exons
not used in building the gene models are presented on the web pages
associated with F9H3. Genes are numbered according to the scheme
BAC.gene.number. Typically, these numbers progress from 1 upwards
as one moves from position 1 of the BAC. Protein sequences encoded
by the genes are assigned to a functional category with the aid of
similarity searches and comparison to the Prosite
(<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam
(<http://pfam.wustl.edu/>) libraries. A description of these
categories can be found at
<http://antjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are
typically located by TBLASTX analysis and an attempt is made to
classify the function of each repeat as either transposon, putative
microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence
concerning the annotation of F9H3, please direct email to Larry
Parnell at parnell@cshl.org.
BAC F9H3 was sequenced as part of the arabidopsis genome sequencing
effort of the Cold Spring Harbor Consortium. For
additional information, please see <http://www.cshl.org/arabweb>.
Fingerprint data indicate F9H3 overlaps with F4C21 toward the
telomere and with T5L23 toward the centromere. F9H3 contains marker
T1233 at 18.8 cm on the Lister & Dean RI map.

Location/Qualifiers
1..95519
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"

FEATURES
SOURCE

```

/misc_feature
1..31783
AC005142, from position 26 to 31806
851..1719
/gene="F9H3.20"
/pseudo
/evidence=not_experimental
complement(981..1120)
/notes="function=putative_transposon; similar to F6K23, GenBank accession number AF058919"
accession number AF058919"
/rpt_type=dispersed
3173..3969
/notes="function=putative_transposon; similar to F6A1, GenBank accession number 297336; moderately similar to reverse transcriptase gene of Dk15 non-LTR retrotransposon, GenBank accession number J47193"
/rpt_type=dispersed
5885..6303
/notes="function=chloroplast_repeat"
/rpt_type=dispersed
complement(5653..12091)
/gene="F9H3.19"
/notes="encodes putative calcium channel, genomic copy of EST W43766; genomic copy of EST T43251; EST T43251 apparently encodes intronic sequence; possibly alternate exon 1 from position 12772 to 12845 is predicted by a NetPlantGene splice donor but has no EST coverage; gene model last edited on 27 Jan 99"
/evidence=not_experimental
complement(join(6553..5746,6833..5870,6991..7062,7175..7261,7436..7526,7682..7745,7985..8153,8229..8380,8487,8757..8814,8891..9034,9127..9192,9311..9383,9530..9591,9745..9822,9929..10034,10127..10293,10331..10451,10557..10580,10644..10774,10850..10930,11292..11403,11500..11643,11972..12091))
/gene="F9H3.15"
/notes="functional catalog ID=07.0"
/codon_start=1
/evidence=not_experimental
/product="putative calcium channel"
/protein_id="AAD11598.1"
/db_xref="GI:4206210"
/translation="MSLRHTVPQTGRPLAAEAVGLEEPYNIIPVNNLADHPSLR
FQVRAAAALKTVGLDRPPYVQWBSHYDLDDMLALFTFCOKDNVNRQEHVHLA
NAQRLSPDDNIDSLDSAVRRFRKLLANYSWCSYLGKSNINISNDPNSREL
JYVGLYLWGEAANLRFWPECIFYIFHNMASELNKLDCLENDENTGQVPLPSLGEN
AFLTVGKPIYDTIOAEIDESKNGTVAHCKWRNYDDINEYFWTDFGSKLKPDLGSG
NFKSGKSGVGTQVERRTFFLYRSFRLWMLALFLQAAIIVAFWEKPIOTSSVTR
OLWNLKARDVOVRLTLVFTLWSGMLOALDAASQYPLVSRKTHFRFRLMKVIA
AAWIVAFVLTVINWKKRQDROHNSAATTKIYQFLYAVGAPLVPEI.LALFLIFW
ARNFLEETNMKIFPALTMPFGKSFVGRGLVDNIKISTFWIPLAKFTFSYFL
QVKMIPKSLKMLNLDKDYENHGYGDSNRSFVAILPLPVLVLIQWAIYSS
IVGAVVGLFDHGEIDMGOLRLRFQFAFQINLMPPEQLNARGFNKFGDIHR
LKRVYGRPPFKKLESNOVEANKFALINWEIILAFREEDIVSDREVELLEPKNSDV
TVTRWPCFELICNELLALSOARELIDAPDKWLHKICKNEYRCVAVDASDKHL.L
SIKKVDTREHSIIVFQIINOSIQSEOPTKTFVLLPKIYETIAKLVGLNDESD
SGRVNVLQSLYEIATROFFIEKTEQLSNEGTLTPDASKLLFQNAIRLCPDASNEC
FYFVRNRLHTILSRDSMSPVYNLEARRIAFFSNSLFMNMHPAPQVKKMAFVIT
FYSEVVSKEQLRNTEDEG-STLYILOTIYADEWKNFKERMHREGIKTDSLELTKK
LRDLKASVGGOTLARTVGMYYRALKMLAFDLSASEMDIREAQEIGSVRNQOG
ELGGQSDGVSENDSSRLSASSVSTLYKGHEYGALMKFTYVVAQIVGSOAKKE
POAEIILYLMKQNEALRIAYVDEVPAGRGETDIYSLVKYDHOLEVEYIFPVKLPQ
VKLGEKPNQNHAMIFTRGDVQTDIDMODSYFEALAMRNLLQYENHYHGRKETI
LVGREHIFTGSVSSLAWMSAQETSEVTLGQRYLANPLKVRMHYHPDVFQWFLSR
GGISKASRVINISIDITAFGNCTLRGNVTHHEYIQVGRDVGGLNIGSFQAKVSG
NGEQLSRDVIYRUGHLRDFPRLSFYTVGFPTFMVILTYAVLGRVYLALSGV
EKSALADSTDNAALGVILNQOFILOGLFTALPMIVENSLEEGF.LATWNEIRMOIQ
LSAVFYTFSGMTRAHYFGRTILHGKAKYRATGRGVVEHKGTENTRLRSHFVKAI
ELMLLIVASHPSIARQSLIYIYAMTISNFWLSWIMAFVFPNPSGFWMLKTYVDFE
DFNMWYOGRIKSTSEQSWERKMYEEDHLRNTKAGLFEVILVLFREFFQYGVY
QLKANGSLFVYLFSEWIFYFAIFVLVQIYARDYSAKAIHRYLVQFVLIVLAI
LVIVALLETFHFSFDITSLFAIFPTGIGILLIAOTQRKWKNTY.FNNAVVSVARM
YDILFGILMVPVAFILSWMPGFGSMOTRILFNEASRGLRIMQIVTGKSKGDV"
complement(20861..22302)
/gene="F9H3.17"
/notes="encodes hypothetical protein; gene model last
edited on 5 Jan 99"
/evidence=not_experimental
complement(join(20861..21094,21937..22035,22145..22302));
/gene="F9H3.17"
/notes="similar to F21B7.5, GenBank accession number
AC002560; functional catalog ID=99"
Query Match 29.2%; Score 497; DB 8; Length 95519;
Best Local Similarity 92.5%; Pred. No. 5, 9e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;
Cy 1 TGGGTTTATGGATACATGACAAATATTTATTATTCATGAGTTTATTCGGAAG 60
Db 23303 TCAGTTTATGGATACATGACAAATATTTATATATTCATGAGTTTATTCGGAAG 23362
Cy 61 CATGACAAATATTAATATATCAGTGTTAATAACATGTTTGTTCCTTAAATACATGCAT 120
Db 23363 CATGACAAATATTAATATATCAGTGTTAATAACATGTTTGTTCCTTAAATACATGCAT 23422
Cy 121 TAAAAACAGACATTTGTTTAAAAATCAAAATCTAATCTCTTATATCAACAGCATTGAC 180
Db 23423 TAAAAACAGACATTTGTTTAAAAATCAAAATATATCTCTTATATCAACAGCATTCAT 23482
Cy 161 GAAAAATCAGCTTAAAAAGACAAATTAAGNATGAGATAGAGAGATTTCTTATGAAAA 240
Db 23483 GAAAAATCAGCTTAAAAAGACAAATTAAGNATGAGATAGAGAGATTTCTTATGAAAA 23542
Cy 241 AGAAGACAGACATGTAGGTGAACAAAAATAAGACAGATATCATGATATATTTATGAGAG 300
Db 23543 AGAAGACAGACATGTAGGTGAACAAAAATAAGACAGATATCATGATATATTTATGAGAG 23602
Cy 301 GTGIGTAGATTTATTTTAGGAGGGGAGAGAAATAGAAAAGAAAATGACATGGTGAA 360
Db 23603 GTGIGTAGA-TATTCAGGAGGGGAGAGAAATAGAAAAGAAAATGACATGGTGAA 23662;
Cy 361 TCTGAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAGAACTCCATCGGTAAAGTC 420
Db 23662 TTTCAAGAAGATGAATTTGTGTTAAAGATGAAGAGAGAGAACTTCATGGCTAAAGTT 23721

```



```
/note="function terminal_flanking_direct_repeat: cognate B
of flanking direct repeat of IIR transposon from position
5:45 to 6873; cognate A resides from position 5:45 to
5:49"
repeat_region
/rpt_type=flanking
/complement(7885..7924)
/note="function:putative_microsatellite"
/rpt_type=tandem
/rpt_unit=AT
/complement(7905..8519)
/note="function-unclassified; similar to MX110. GenBank
accession number AB005248"
repeat_region
/rpt_type=dispersed
/9482..10923
/note="TSL23.3"
/note="encodes hypothetical protein; identical to F9H3.17,
GenBank accession number AF071527; gene model last edited
on 5 Jan 99"
CDS
/translation="not experimental"
/9482..9639,9745..9847,10690..10923
/note="TSL23.3"
/note="identical to F9H3.17, GenBank accession number
AF071527; similar to F2187.5, GenBank accession number
AC002560; functional catalog ID=99"
/codon_start=1
/evidence=not experimental
/product="hypothetical protein"
/protein_id="AAB15310.1"
/db_xref="GI:4263041"
/translation="MVKLTKRGGLVRLAAGAAALIVMTSRERASEFLA:SLEA
KYDMAAKFYVIANVSVSFLVLELPKESLWKEVVDLDVMTLCTSS:GAALA
VAQCKKGNANAGWPIGQVPKFCQITGAL:AGEVALVLYVLLLSLRAVDVPL
LQKS"
12953..12635
/note="TSL23.4"
/note="encodes putative glucan synthase component;
identical to F9H3.18, GenBank accession number AF071527;
gene model last edited on 5 Jan 99"
/evidence=not experimental
/codon_start=1
/product="putative glucan synthase component"
/protein_id="AAB15311.1"
/db_xref="GI:4263042"
/translation="MSLRHTVPPQIGPLAAEAVGIEEPYNI:PVANLLADHPSIR
FEVPAARALXIVGDLREPPVQVRSHYDCLMLALFGQKQNVKRCHEMV:RIA
NAQMLSPPDN:DLSDNVVRRFRKLLANYSKCSVLGKSK:WISDRNDSREPL
LVGLFLLWGSAALNRPEPCTICITFHMASLNKIL:ECUDENIGQYPLPS:SGEN
AFLTGKVPYDIIOADESKNGIVAHCKWKNYDINEYF:DRCFKJRWPDJLGS
NFFKSKSGVGHVGTERTFFLYSRFDRLWYMLALFLQAAIIVAMEKPKDLSVY
QJWNLKAKGQVQVLLTVLTWSGKRLQAVLDAASQPLVSKR:KNRIFRMLKVA
RAVVAIVTVLNTWKOKROSKNSNAATKIYQFLAVAGFLVLCALALF:TPW
MKNPLETKNIFPALTWPFQKSEVGRKREGLVNLIKYSTFIYVLAFTKFT:SYFL
QKMPKLSKLLWNLDLWVQKREYFGDSNRFSVAL:MLPVY:ZILMDIQ:WFA:YSS
IVGAVYDFHLGEIR:RMQCK:RLRTOFFASATQFNLMPPEQLNARGTKNFK:HR
LKLYRGFGRPKLESNQYAKKFLIMNEILAFREEDIVSDREVELLE:PKNSWDY
TVIRNPFCLNELLALSOARELIDAPQKMLWIKCKNYRRCVAVYAEYSIKILL
SIIKVDEBSLITVFQIINOSIOSEQFTKTRVDLIPKIYETLCKLVG:VNDRETD
SGRVNVLQSLVEIATQRFIEKKTEQSLNEGLTPRDPASKLLQFNAR:LDASNE
FYQVRRLITLSDMSHSVPVNDERRRIAFFNSLPMNPHAPQVKGNMAF:SVLT
PYSEVTVYSKQIURNETEDGISTLYLQTIYADENKPKERMHEGKIDSELMTIK
LDRLRLWASRGQTLARTYRGWMMYVRL:KMLAF:LDSEMDIRSGAGVSGVRLQ
ELGGQSDGVSNDRSSLSRASSYSYLYKHEGYTALMKFTYVACQIYGSQAKKE
PQAEIILMKQNEALRIAYDEVDPAGRGETDYYSVLVYKVDHOLEKEVEIFRVKLPGP
```

Query Match

Best Local Similarity

Matches 545; Conservative

29.2%; Score 497; DB 8; Length 116448;

92.5%; Pred. No. 5.6e-77;

0; Mismatches 40; Indels 4; Gaps 2;

```
QY 1 TGGG:TTTATTCGATACATCGACAAATATTATTATTTCATGAGTTTATTTCGATAG 60
11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16 248: TGAATTTTATTTGATGAACATGACAAATATTATATATTATCATGAATTTTGTG 8422
17 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 57 CATGACAAATATTAAATATATCATGTTTAAATAACATGTTTGTCTTAAATACATG 120
18 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
19 842 CATGACAAATATTAAATATATCATGTTTAAATAACATGTTTGTCTTAAATACATG 8362
20 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 TTAATAACAGACATTTTGTAAATCAAAATCTTAATCTCTTATATCAACAGCATG 180
21 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
26 8361 TTAATAACAGACATTTTGTAAATCAAAATCTTAATCTCTTATATCAACAGCATG 8352
27 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GGAATATTCAGGTAAACAGACAAATTAAGATGAGATAGAGAGATTTCTA:GGAAA 240
28 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29 831 GAAATATTCAGGTAAACAGACAAATTAAGATGAGATAGAGAGATTTCTATG 8242
30 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 AGAAGAGAGACATGTAGGTGAACAAATTAAGAGATATGATGATATATTTTAIGAG 300
31 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
36 8241 ACAAGAGAGACATGTGGTGAACAAATTAAGAGATATGATGATATTTTAIGAG 8182
37 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 331 GTGGTGAAGATTTATTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATG 360
38 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 8131 GTGGTGAAGATTTATTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATG 8123
40 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 TCTGAGAGAGATGAATTTGTTTAAGATGAAGAGAGAGAGAACTCATGCTTAAATC 420
41 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46 8122 TTGAAGAGAGATGAATTTGTTTAAGATGAAGAGAGAGAGAACTCATGCTTAAAT 8063
47 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 TCGTAAACAGATGAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
48 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 8062 TCGTAAACAGATGAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8006
50 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 CTAACTATTGCCAAATTTCTGTAGCCGCAAAATCTACTATTGGTCCAGGTTTAT 540
51 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 8065 CTAACTATTGCCAAATTTCTGTAGCCGCAAAATCTACTATTGGTCCAGGTTTAT 7946
57 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 TATTCCTTTGAAGTCAAAAGTTATCTTACATCTACTCTTAAATATA 589
58 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 7945 TATTCCTTTGAAGTCAAAAGTTATCTTACATCTACTCTTAAATATA 7897
60 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 10
ATCHKV9/C
LOCUS ATCHKV9 159629 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 9.
ACCESSION AL151497
VERSION AL151497.2 GI:7270670
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi-
s. (bases 1 to 53060)
REFERENCE
1. Spegele, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M.,
Vil, D.M., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Shekher, M., Schutz, K.K., See, L.H., Swaby, I.,
Hadermann, K., Dedhia, N., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 21275 to 137726)
AUTHORS Zhong, J., Ma, P., Parnell, L.D., Chen, C.N., Chen, E.Y., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 3 (bases 137729 to 137730)
AUTHORS Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 159629)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
```


lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@jbsrc.ac.uk
 information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/prc/jthal/>
 this fragment has an overlap with ATCHRIV8 at the 5' end and an
 overlap with ATCHRIV10 at the 3' end.

FEATURES

Source

1. 159529

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

6741..9588

/gene="AT4g03480"

complement(6741..7064,7161..7708,7783..8101,

8409..8866,9283..9343,9445..9588);

/gene="AT4g03480"

complement(join(6741..7064,7161..7708,7783..8101,

8409..8866,9283..9343,9445..9588);)

/gene="AT4g03480"

/note="contains similarity to mammalian ankyrins

similarity to similar to A. thaliana hypothetical protein

F21B7.8, GenBank accession number AC002560

Contains ATP/GTP-binding site motif A (P-loop) AA297-304

contains EST gb:37081, 742092"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAB77834.1"

/db_xref="GI:7270671"

/translation="MYRSHLRATPEASPLKPASSPNCRLSLSPVELRVPFIRDEBP
 KYVDIFKSYONSPELPCCDKSREASLSRESLENLGEVYRNNPFIISMRAQNI
 ELLEKLKSVETMAQLKSDGGSVHLAASGHELVNITECHLLEEFNSKTCIP
 LHVAAAGASVAVKALVASVYFSPRVPEEDRLNIYVKJIDGETPLHAAIKDIRE
 KAEERIKLSLSLTHMWRSCISFSDASTQMETAAALVNADQHASFLANKDKTSP
 IYLVAVAGNLSVLVAMLRPNKIQGKSTLASQLEGRKSLLEAKAKAKTDVAVIL
 NDPLSVNERDEEGRCLSVASGMVYKIGCKLLDRSTKSVYCEPKDGSFPIHMAVEK
 GHLKVVKETLKRCPDSKELVNGQNMHLIAKSAKVSFLLSYIRLDEINELLEEQ
 DYDGNAPLHATINRCRTVDLAFASTETKILNIQKDGURPDIAEJNLQPDYVL
 RERLTMLVLCVYAKSGWLPSTGTMHSRSEPLDARKYKDIKAILLVATLVATVAT
 FAAGTIPGCFNSSAPNMGMATLADDSTLFFTVLDTLAWSSIVAIVAL."

complement(6741..7064)

/gene="AT4g03480"

/number=1

complement(7065..7160)

/gene="AT4g03480"

/number=1

complement(7161..7708)

/gene="AT4g03480"

/number=2

complement(7709..7782)

/gene="AT4g03480"

/number=2

complement(7783..8101)

/gene="AT4g03480"

/number=3

complement(8102..8408)

/gene="AT4g03480"

/number=3

complement(8409..8866)

/gene="AT4g03480"

/number=4

complement(8867..9282)

/gene="AT4g03480"

/number=4

complement(9283..9343)

/gene="AT4g03480"

/number=5

complement(9344..9444)

/gene="AT4g03480"

/number=5

exon

complement(9445..9588)

/gene="AT4g03480"

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

/number=6

```
/protein_id="CAB77835.1"
/db_xref="GI:7270673"
/translation="MKNESRTSEISKGIPIESGVGSGREDYLDORVAFNHCAI2R
NNVSTPSLDLSTLDESEKPKDPKTKMAAVRAKKNYLRNNYSYISVAPILVNDKQ
NITLHLAASSSHVSLVRNIIIOKCPGLLLKSNMGEVALHLAAEAGHLDVWNLLDFIN
DISCNLPAVKRIYFAKNKQDTHLVAKGKHEVASYLVSAAKS:SPVANRGGSPF
LYLAIEAGHTSLVMTCHGTNELSSKVGCRSIVHAALKANKRDIALLSKDASLIN:
RDEGRTS:SFGAS:GYVGGFSVLTDRKRVVVSDDUGLPTTHAAKYGHVQI:LEEL:
KHCPEAIELJRDQGNILHLAAKYGLKVIKELSCCKDKKKKLLINEQDNGNTPLEH
LAPNSHPKVVSMFTWDRVLDKRNZYGFALDVAEENIDSSYIVHQRFLMALINP
GATSKSTPITENSKRPKDPGKYKORVNTLMELVATMTFAGETLFGYKNSFP
HGNVLAKRATFAYVCDTILAYSS:ITIVALLIMAC:GELS:LLKAFNIALPFLGL
ALTSMAIFMAGTYVAVSHLPLLYGVFLGIGITFLLVLLVLLVYVSPYAHQAQ:LLKE
:FYYPYF:KLLAAGDNKNIVDYVYASDK"
exon 14170..14204
/gene="AT4903500"
/number=1
intron 14205..15102
/gene="AT4903500"
/number=1
exon 15103..15848
/gene="AT4903500"
intron 15849..15938
/gene="AT4903500"
```

```
Query Match 29.2% Score 497; DB 8; Length 159629;
Best Local Similarity 92.5%; Pred. No. 5.3e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

QY 1 TCGGTTTATTGGATAACATGACAAATATTATTATTTTCATGATGATTTTATGGSATAG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25758 TGAGTTTATTGGATAACATGACAAATATTATATATACATGATTTTCTTGATAG 25659

QY 61 CATGACAATATATATATACATGTTTATACATGTTTGTCTTAAATACATGCATT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29698 CATGACAATATATATATACATGTTTATACATGTTTGTCTTAAATACATGCATT 29639

QY 121 TTAATATCAGACATTTGTTTAAATCAATCTATCTCTTATATACAAACACATTCAC 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29638 TTAATATCAGACATTTGTTTAAATCAATATAATCTCTTATATACAAACACATTCAT 29579

QY 181 GGAATATTCAGGTAAAGAGAGAAATTAAGAATGAGAGATAGAGACATTTCTATGSAAA 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29578 GAAGATTCAGCAAAAGAGAGAAACAAAGATGAAGATAGAGACATTTCTATGSAAA 29519

QY 241 AGAAGAGAGACATCTAGTGAACAAATTAAGAGATATGATATATTTTATGACAG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29518 AGAAGAGAGACATCTGGTGAACAAATTAAGAGATATGATATATTTATGACAG 29459

QY 301 GTGGTCAAGATTAITTTAGAGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGAA 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29458 GTGGTCAAGA-TAICTAGAGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGAA 29406

QY 361 TCTGAAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAAATGCTATGCTTAAATTC 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29399 TTTGAAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAAATGCTATGCTTAAAGTT 29340

QY 421 TCGTAAAGAGATGAAGAGAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29339 TCGTAAAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 29283

QY 481 CTAACTATTGCCAAAATTTCTGTAGCGGCAATATCTATTGTGTCACAGTTATTGTG 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29282 CTAAATATTGCCAAAATTTCTGTAGCAGACAGTACTATTGTGTCACAAAGTTA:TTTGT 29223

QY 541 TATTCTTTGAAGTCAAAAGTTATTTCCTTACATATCTCTTAAATATA 549
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29222 TATTGTTTCAAGTCAAGGTTATTTTGTATATATATCTCTATATATA 29174
```

RESULT 11
ATH250185

```
LOCUS ATH250185 369 bp mRNA linear PLN 15-AUG-2001
DEFINITION Arabidopsis thaliana mRNA for NIMIN-2 protein (nimin-2 gene).
ACCESSION AJ250185
VERSION AJ250185.1 GI:12057155
KEYWORDS nimin-2 gene; NIMIN-2 protein.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 369)
REFERENCE 1
AUTHORS Weigel,R., Bauscher,C., Pfitzner,A.J. and Pfitzner,U.M.
TITLE NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins
from Arabidopsis that interact with NPR1/NIM1, a key regulator of
systemic acquired resistance in plants
JOURNAL Plant Mol. Biol. 46 (2), 143-160 (2001)
MEDLINE 21335044
PMID 11442051
REFERENCE 2 (bases 1 to 369)
AUTHORS Pfitzner,U.M.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Pfitzner U.M., General Virology, Institute
of Genetics, Emil Wolff Str. 14, 70599 Stuttgart, GERMANY
COMMENT Related sequence: AB023041..
FEATURES
Location/Qualifiers
1..369
/organism="Arabidopsis thaliana"
/cultivar="Col-0"
/db_xref="taxon:3702"
/tissue_type="green vegetative tissue"
/dev_stage="3-wk-old"
1..369
/gene="nimin-2"
1..369
/gene="nimin-2"
/function="putative role in SAR"
/codon_start=1
/evidence="experimental"
/product="NIMIN-2 protein"
/protein_id="CAC19845.1"
/db_xref="GI:12057156"
/db_xref="SPTREMBL:Q9LUA3"
/translation="MNNSLKKEERVEDNGKSDGNRGKIPSTEVVTVTEEEVEDEFFKI
LRVHVATRTVAKVNGVAVGELPSSKKRKRSQNLGRNSLDCNVRDGEDEINRVGI.
OGLGLDLNCKPEPDSVLSL"
BASE COUNT 113 a 53 c 121 g 82 t
ORIGIN
Query Match 21.7%; Score 369; DB 8; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.5e-54;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 863 ATGAACAACACTCTTTTGAAGAAAGAGAGACCGCTAGAGAAGATAACGGAAATCTGACGGT 922
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGAACAACACTCTTTTGAAGAAAGAGAGACCGCTAGAGAAGATAACGGAAATCTGACGGT 60

QY 923 AACAGAGGAAACCGCTGACGGAAGTCTTCGCGCGGTAAACGGAGAGAGTGGATCAG 982
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 AACAGAGGAAACCGCTGACGGAAGTCTTCGCGCGGTAAACGGAGAGAGTGGATCAG 120

QY 983 TTTTTCAGATATTACGGAGAGTACACGTGGCGACACGACGCGTTGCGAAAGTTAAACGC 1042
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TTTTTCAGATATTACGGAGAGTACACGTGGCGACACGACGCGTTGCGAAAGTTAAACGC 180

QY 1043 GGTGTTGCTGAGGAGAGATTACCGTCTTAAGAAGAGAGAAACCGAGTACAGATCTTGGGTTG 1102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 GGTGTTGCTGAGGAGAGATTACCGTCTTAAGAAGAGAGAAACCGAGTACAGATCTTGGGTTG 240

QY 1153 ACAAACTCGTTGGATTGTAACGGCGTTCGACACGAGAAATTCGATGAGATTAAATCGSGTC 1162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AAAAACTCGTTGGATTGTAACGGCGTTCGACACGAGAAATTCGATGAGATTAAATCGSGTC 300

QY 1163 CGGTTACAGGGTTTGGGTTTGGATCTGAACGTAAACCGGAACACAGACGCGTTAGTITA 1222
```

Db 301 GCGTACAGGTTGGTTGGATCGACCTGTAACCGGACACAGCGGTAGTTTA 360
 QY 1223 TCGTGTAG 1231
 Db 361 TCGTGTAG 369

RESULT 12
 AC007203/c
 LOCUS
 DEFINITION Arabidopsis thaliana chromosome I BAC T10P12 genomic sequence, complete sequence.
 AC007203
 AC007203.3 GI:4956903
 VERSION 57G.
 KEYWORDS Arabidopsis thaliana
 SOURCE Arabidopsis thaliana
 ORGANISM Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots; Spermatophyta; Magnoliophyta; Streptophyta; Embryophyta; Tracheophyta; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 95190)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenevskaja,I., Kim,C., Lenz,C., Li,J., Liu,S., Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vrotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Unpublished
 2 (bases 1 to 95190)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenevskaja,I., Kim,C., Lenz,C., Li,J., Liu,S., Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vrotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Direct Submission
 Submitted (03-APR-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 3 (bases 1 to 95190)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenevskaja,I., Kim,C., Lenz,C., Li,J., Liu,S., Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vrotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Direct Submission
 Submitted (04-JUN-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 4 (bases 1 to 95190)
 Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenevskaja,I., Kim,C., Lenz,C., Li,J., Liu,S., Lueros,S., Schwartz,J., Shinn,P., Toriumi,M., Vrotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 Direct Submission
 Submitted (16-JUN-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
 On Jun 4, 1999 this sequence version replaced gi:4678167.
 Bases 1-6718 of clone T10P12 overlap with bases 122,554-129,271 of 'IGF' BAC clone F1121, gb|AC005687
 e-mail for correspondence: arabseq@stanford.edu
 Genes with similarity to proteins in the databases are described as 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCAN.html), Rexa (V.Solov'ev & A.Salamov, Sanger Centre, http://genome.c.sanger.ac.uk/), NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/NetPlantGene.html) and eMotif(Nevill-Manning, C.G. Wu, T.D. S. Brutlag, D.L., http://motif.stanford.edu/projects.html).
 Location/Qualifiers
 1. 95190
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="T10P12"
 1. 6718
 /note="Overlap with bases 122,554-129,271 of 'IGF' BAC clone F1121, gb|AC005687"
 9474..10528
 complement(19003..23764)
 /gene="T10P12.1"
 21074..21678,21933..22099,22247..22640,22931..23131,23362..23764)
 /gene="T10P12.1"
 /note="Similar to reverse transcriptase; Similar to reverse transcriptase, part of putative retrotransposon"
 /codon_start=1
 /protein_id="AAD39270.1"
 /db_xref="GI:5080760"
 /translation="MDMIFNVGDKVRRKTIENSKTAAEAWGILDIYLVKSLFNRVVLQKLYNYRMNSKLTLENVDEFQKMSIDLANLQIQVDEQVQAVQAILLSALFSDCM LKFTREGGIRKLDVSAISKELESDSTSEVNLGSGESQKQHGIEWIDTCEPHM TPKEHYHIDFEAKSGKVRPNNSFLVKGIGKVRFTNQDGTSCGNDRIKKECNF VVGTTRTSFTAQHVKELDYVHSDLWGSNNVPSLSKCYFIPTDWSKRVNIN FLNTKDAFTFTWKMKVETOSERKLKRTDNGLESCNCKFQGFYKKEIYVRHTS PSSAVENK1PEELWTSVIPSLEGLRRFGCIYVHSGEKLDKPKAKGVGVSPGHVD TESENQRTQIADRPGRVILPSRLKDYEMDEVIDETAGYATETEZGGNPEAS PQEALQSDSGKWLAEDEIQSLKNTVVERNKSQRPICGKVVFRKKAGIAYE KPRKARLVYKSAQKEGIYFISPVVKVYSIRLLSIVFHLMDLQMDKPTAFLE HGYLDIMFDQPEYVHEKPYRIYLLYVDLITASKDKHVKCEKLVLLSETEM KQLDADKILGMEIIRDEKGTITISQGYLLKLRPATYEVQKQKVMKIVNYSRSG LSLMYSMIGTPDLAHSVGLLGMKRPKLDHWQAVKWLICVIGTMDRKLCTKNSG TNEIWLKGLVNLGELFTQSRVNHCDQSQVIALAKNAVYHRTKHUVKHFTEWAK ANSLDKLGGSGLEEEKVKQVLAWKKEYDDAEATYKLVVLEKELKIPVFNFLIP 29564..30696
 complement(31637..35319)
 /gene="T10P12.2"
 34104..34307,34383..34539,34828..35319)
 /gene="T10P12.2"
 /note="Similar to reverse-transcriptases; protein is flanked by LTRs, potential retrotransposon"
 /codon_start=1
 /protein_id="AAD39271.1"
 /db_xref="GI:5080761"
 /translation="MMEQQLRIGTGYFGSGTSPEEADSWRSRVERNSGSKSCSEVYL DLAHYLEGDAHLWRSVTAKRQDDFWADFEAEFNKCYPOEADRMAREFLEITO GERSVYFDEFNRLLYAGRMEDQAMRFLRLGRPDWRVRCRVSQVATKAALVE TAELMVAVVQPAVQOQAVQOQVQOOLAHIGAPOGYTTREIGSTNRATIGL LAHEVCVTVLVQGVAEHLVDSGASHCFTPKASRGNTRGDQFGQSVKVGAGOF LEVLGRKATVEAKNIENGCEATYLSMSPVQVAVSNIRVCEFDVFSIQGLP PWSMDPTTIVLEPGTALLSKHPTQLOORWIDELLQLRGATCFSKDLTSGYHQIPI AEADIRKTRTRYHFHYVIFPARAELDAIWIVDLRTKSAHLAIRKTDGAVALI KVVSEYVTKHGVPSVSDRSKFTSAFWRAQAEKTKVQMSIAYHPOTDQSSERT IOTLEDMRCLVDGWHADQLSLIEFANNYQASIGMTPEVLYRGPCRTQIQSWT QVGRSIVGANYVOETTERIGVLKLNKKEVDHRSVADKRRELEFEFGERVYLKMA MLRGNRSISETKLSPRYMGPFPRIVERVVPVAYLELTDVRAFKVHFVHMLKCLH KDDVLAKEIKEDLPQNMNTELRGVRRLRRRIKELRRKKIPLIKVLMDCDGVETWEPEVRRKARF"
 35970..37136
 37249..42330
 /gene="T10P12.3"
 join(37249..37347,37714..38098,38266..38532,38897..3909",39215..39508,39603..40889,41069..42330)
 /gene="T10P12.3"
 repeat_region
 score
 CDS
 repeat_region
 score
 CDS

/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5030742"
/translation="MAVNGAKGMDKDRKDPFKKEKEEDSDGDFVGNFELCLCAKA
CHEEKLENDRIYRVSIVSTLERTEIRAGKASEPHDFHQSPKVGSSSIHSRRAPMECE
RYLNDLPDNCYGLKLTIEPMYGLIITNGFQWRSEFRVRYRHNDUKJNLVSJLS
TEFEDLSITQVGLDVSHLEKIFVYNGLKPMEKVIKMKESRGLPQKVAVIRMESSGAY
OMYSEKGNVAAKATYKNSYKPPILITSGAPVAHRSKLKLCADKNLEVLNGLCVGSH
CSVQGVKFLMGLGVVEEGNGESLGLGVGVVGLGFWLESLAVPKLVEHESDLTEGS
LSUSLATEMLPVEVVEUSEPTTIDHIELPFPFGLHSLSLAPGVSSILVRGIRY
PHATNVLEKMAEVMVAVTGIIRPTTIDPKPPIVATFVLELLDELIGAMFTSKIDLCSSYH
QIRMKPEDEIKTAFTIECHYESEFLVMSGLFNATPATFQSSMMRIKSKLKLKFLVFFQ
HILIVSPVSEVHKHREVLQDLADKNLKNMKGSEGVQGVVEYLGHIISANGVVTQS
IKTEAMENKMPVSNLKLRSGLFTGTYGYIKFVRYGSIKATLTLELLKRXKXWLTNQC
DAEYVQKQMLAPVGLGFQDFKRVFLEDETVGTGTVGAVLMQEKRETYFSEHLETERE
QLASVREMAIVMLVKMKHNYLQKRVKHIDQSLAELEQGVNMEDEIRGNIRK
IGCDMOTIVYGVANKAAGDSLRQHEVVAIILALATPYVVIQLODLYKEIDEETHIQS
VIKIHNDLQGGSHYQJVNRIWYKJLQVFPKRASSH IPLIQLGYEDTVMEDISVDFTE
GLPTSGGVNIIIVVORLSKAVYFLGLKHLFRAIDVANKFVSEVVKJYGFPSKISVSDR
DRIFLAGAVKMDKELSGTOLKHSITSYRPTDGGTEVLRNCLETYLRKCTFSEHURAWHK
YLAVERSEINTAYHSAKMDKPPRYVGVGRDPSLVRFELGSIENWDDEVOLKREJMLR
HIRANQAGEIEMKRNALKTRFVAVGDVHLKQPIQRRTVVRVYQKILAAKFPFG
PFVEIVERGIVYRIINLPKSSITHNVYSQLKKEGVQGVQVATNPKKILVDEIDFP
EKTVISFTDKGNEFKELVQKQDGRVSENSMWPKEFVQVQFQKLEDDILFARSDIM
IHSAVYRQRTGPRQHLRVEGEDVAVTPPSLTIRNSAKDPGFQVLGQSFSSCHLPTRC
TA"
complement(47812..49231)
/gene="Tl0p12.4"
CDS
complement(join(47812..47877,48046..48223,48314..48435,
48706..48828,49010..49261))
/gene="Tl0p12.4"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39273.1"
/db_xref="GI:5080763"
/translation="MAISLATAIVISCPITPSSNSASPSKTSSVRLPSCQIRRFQSVQ
SPSSITFAPLTVSSPTNVLFPFADIDRETKENFVLVLTSSPMVNEUKPLADGLAN
YAPSLPLGLHLSERPDVSSILLADVAGVRAAKKQIFNSFDLLONSUKPLVDFEYA
TWGPGCQWPTILNEVSETUKDIAVVKDITKPISLANKYQIATPILTFIDRUKLW
DRFEGALPANOLVERENSLQVKQ"
complement(50976..52602)
/gene="Tl0p12.5"
CDS
complement(join(50976..51522,51912..52404,52575..52602))
/gene="Tl0p12.5"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39277.1"
/db_xref="GI:5080767"
/translation="MCLAOMTINSQLPVRVILGFLMTKAWTAHDXLPLLEIKRKSS
WTGAFSYGCRLOSLIKSVLMSITNFNSVAVGKLRKMEKEELCSAFINSGPQDKPH
AKIAWAKPLKPGGGUGLAPLKEITVCGKKIWRILASQISLWQGVWVTLFIPNN
FWATKASYGGSQWNCWVQATPKFATITWLGMMHRLSILDRCKMKQCASTCFVQO
PLETRHLLPFCHYANQIMPIKAGPMGVQVTSNMQISLADTASLACTSIPEPIGFFVSM
FKPPTQRESREITVGNKEIQECSQINNRKPESTIQYETESPFSSATREIKMKPTSSVEG
PLEANLSLTENEFLFLKQCT"
complement(60126..62065)
/gene="Tl0p12.6"
CDS
complement(join(60126..60521,60567..60800,60875..61272,
61788..62065))
/gene="Tl0p12.6"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39274.1"

```

Query Match      19.5%  Score 330.8  DB 6;  Length 95190;
Best Local Similarity 78.3%;  Pred. NO. 3.7e-48;
Matches 503;  Conservative 0;  Mismatches 67;  Indels 72;  Gaps 6;

1 TGGGTTTTTATGGCATACATGACAAATATTTATTTTCATGAGTTTATTCGGAAG 60
68585 TGGGTTTTTATGGCATACATGACAAATATTTATTTATTTATTTATTTATTTGTTGGA 68525

```

QY	61	CATGACAAATATTAAATATATCATGTGTTAATAACATGTTTTTGGTCTTAAAAATACATGCATT	120
DB	58525	CAAGCAAATATTAAATATATCAGTGTTTAATAGCATGTTTGTTCCTTAAAATACATGCATT	68456
QY	121	TTAA-TATCAGCATTTGTTT-TAAATCAAAATCAATCTCTTATATCACACGACATGA	179
DB	58465	T-AAGAATCATCATTTGTTT-TAAATCATATCTAAATCTCT--TAICACAACGACAATGA	68408
QY	180	CGAAATTCAGGTAAACAGAGAAATAAAGAAATGACAGATAGAGCATTTCTATGCAAA	239
DB	63407	TGAGAGATTCAGCCAAACAGAGAAAGAGAGATGAGAGATAGAGAGATTTCTATGCAA	68348
QY	240	AAGAACAGAG-GAACATGTAGTGTGAACAAAATAAAGAGATATGATGATATATTTATGAG	298
DB	63347	AAGAAAGATAGGAACATGTAGTGTGAACAAAATAAAGAGATATGATGATATATATCAG	68288
QY	299	AGTGTGTGAAGATTATTTPAGGAGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTG	358
DB	63287	AGTGTGTGAAGA-TATTTAGGAGAGAGAGAGCAAAAGAGAAAGAAATGACATGGTG	68229
QY	359	ANTCTGAAA---GAAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAGAACTCCAIGGCTA	415
DB	63228	AACCTGAAGATGAAGATGACTTGTGTCAAAGATGAAGAGAGAAAGAGAACTCCAIGGCTA	68169
QY	416	AGCTTCGTTAAGAGAGATGAAGAAGAGACAAAGAGAAAGAGAGAAAGAGCAATGCTAAA	475
DB	68168	AGCTCTC-----	68162
QY	476	ATAGACTAATATTGGCCAAAATTTCTAGCCGACAAATACTATTTGGTCCAAAGTTATT	535
DB	68161	-----ATTACC AAAATTTCTACAGACTAGTACTATTTGGTCCAAAGTTATT	68113
QY	536	TTGTCTATNTCTTTTGAAGTCAAAGTTATTCTTTACATATACTCTTAAAAATATAGCCGAT	595
DB	68112	TTGTCTATNTCTTTTGAAGTCAAAGTTATTCTTTATATATACTCTTATTTCTTATATAT	68053
QY	596	ACCAATTTTCCACACATGCATCCCTTTATTCCAAAGTCA	637
DB	68052	ATATATATATATATATATAGAAATTTTTAGAATTA	68011
RESULT 13			
LOCUS	166494	7218 bp DNA linear	PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.		
ACCESSION	166494		
VERSION	166494.1	GI:2724471	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	bases 1 to 7218)		
AUTHORS	Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.		
TITLE	Recombinant foxpox virus		
JOURNAL	Patent: US 5670367-A 14 23-Sep-1997;		
FEATURES	Location/Qualifiers		
source	1..7218		
	/organism="unknown"		
BASE COUNT	1944 a	1491 c	1486 g
ORIGIN	1929 t		
	368 others		
Query Match	6.1%;	Score 103.2;	DB 6; Length 7218;
Host Local Similarity	13.2%;	Pred. No. 1.6e-08;	
Matches	63;	Conservative 240;	Mismatches 173; Indels 0; Gaps 0;
QY	5	TTTTATGGATACATGACAAATATTATTATTTTCATGAGTTTTTATGGATAGCATG	64
DB	1593	TAGATTTTACATACATATATTCTAAATATGAAGTGGTGATCTGTGACTAGCGTAGATC	1534
QY	65	ACAAATATTAAATATCATCGTGTGTAATAACATGTTTTTGTTCTTAAAAATACATGCAATTTAA	124
DB	1533	GCTTCAGACGCACTATTTCAGTTTCAAAAACGCCGATGAGGCATCAGTCACTGTAATACCT	1474

* 34188 50517: contig of 16330 bp in length
 * 50518 55617: gap of 100 bp
 * 50618 65071: contig of 14454 bp in length
 * 65072 65171: gap of 100 bp
 * 65172 81054: contig of 15883 bp in length
 * 81055 81154: gap of 100 bp
 * 81155 95064: contig of 13930 bp in length
 * 95065 95184: gap of 100 bp
 * 95185 133197: contig of 38013 bp in length
 * 133198 133297: gap of 100 bp
 * 133298 171763: contig of 38466 bp in length.

FEATURES

Source
 1..171763
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-358G9"
 /clone_lib="RPCI-24 Male Mouse BAC"

misc_feature
 1..1159
 /note="assembly_fragment"
 misc_feature
 1260..2312
 /note="assembly_fragment"
 misc_feature
 2413..3453
 /note="assembly_fragment"
 misc_feature
 3554..5119
 /note="assembly_fragment"
 misc_feature
 5220..7819
 /note="assembly_fragment"
 misc_feature
 7920..9949
 /note="assembly_fragment"
 misc_feature
 10050..14159
 /note="assembly_fragment"
 misc_feature
 14260..18550
 /note="assembly_fragment"
 misc_feature
 18651..25307
 /note="assembly_fragment"
 misc_feature
 25408..34087
 /note="assembly_fragment"
 misc_feature
 34188..50517
 /note="assembly_fragment"
 misc_feature
 50618..65071
 /note="assembly_fragment"
 misc_feature
 65172..81054
 /note="assembly_fragment"
 misc_feature
 81155..95084
 /note="assembly_fragment"
 misc_feature
 95185..133197
 /note="assembly_fragment"
 misc_feature
 133298..171763
 /note="assembly_fragment"

BASE COUNT 48403 a 34526 c 33542 g 53768 t 1524 others

ORIGIN

Query Match 5.38; Score 90.2; DB 2; Length 171763;
 Best Local Similarity 56.1%; Pred. No. 1.6e-06;
 Matches 170; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
 QY 194 AAAAGAGAAATATAAGATGAGATGATTCATGCAAAAGAAAGAGAGAAC 253
 Db 130691 AGAGAGGGGGGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGA 130692
 QY 254 ATGAGGTGAACAAATAAGATATGATGATATTTTATGAGAGGTGCTGAGAGATTA 313
 Db 130631 AAGAAGAGAAAGAGAGAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGA 130572
 QY 314 TTTTAGAGAGGAGAGAGAAATAGAAAGAAATGATGCTGIGTAATCTGAGAGAGTGS 373
 Db 130571 GAGAAAGA 130512
 QY 374 AATTGTGTTAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 433
 Db 130511 AGA 430452
 QY 434 GAAAGAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493

Db 130451 AGAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 130392
 QY 494 AAA 496
 Db 130391 AAA 130389

RESULT 15

AC125184 280982 bp DNA linear HTG 23-JUN-2002
 Mus musculus chromosome UNK clone RP23-299L9, WORKING DRAFT
 LOCUS
 DEFINITION
 SEQUENCE, 4 unordered pieces.

AC125184
 AC125184.1 Gi:21490707
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 280982)
 McPherson,J.D. and Waterston,R.H.
 TITLE
 The sequence of Mus musculus clone
 JOURNAL
 Unpublished

REFERENCE
 2 (bases 1 to 280982)
 McPherson,J.D. and Waterston,R.H.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park

REFERENCE
 3 (bases 1 to 280982)
 McPherson,J.D. and Waterston,R.H.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submission@watson.wustl.edu
 ----- Project Information -----
 Center project name: N_BA0299L09
 ----- Summary Statistics -----
 Sequencing vector: M13: 0%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 277729 bases at least Q40
 Consensus quality: 279043 bases at least Q30
 Consensus quality: 279926 bases at least Q20
 Insert size: 197000; agarose-fp
 Insert size: 281848; sum-of-contigs
 Quality coverage: 0.00 in Q20 bases; agarose-fp
 Quality coverage: 9.94 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 8161: contig of 8161 bp in length
 * 8261: gap of unknown length
 * 51021: contig of 42760 bp in length
 * 6262
 * 51022 51121: gap of unknown length
 * 51122 84880: contig of 33759 bp in length
 * 84881 84980: gap of unknown length
 * 84981 280982: contig of 196002 bp in length.
 * Location/Qualifiers

FEATURES

```
source      1. 280982
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-299L9"
misc_feature 1..816
/note="assembly_name:Contig10"
misc_feature 8262..51021
/oto="assembly_name:Contig11"
misc_feature 51122..84880
/note="assembly_name:Contig12"
misc_feature 84981..280982
/note="assembly_name:Contig13"
BASE COUNT 90913 a 51036 c 51472 g 87256 t 305 others
ORIGIN
Query Match      5.2%; Score 89.2; DB 2; Length 280982;
Best Local Similarity 54.6%; Pred. No. 2.2e-06;
Matches 178; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 138 TTTTAAATCAATCTAACTCTTATATCACAACGACATTTGACGGNAATTCAGGGTAAAA 197
Db 162899 TTTTAACTGAAAAGAGCAACCTCATCAAAAAGAGAGAAAGCAGAAAGAGAGAGAGA 162840
QY 198 AGAGAAAATAAGAGATGAGAGATAGAGATTTCTATGCAAAAAGAGAGAGACATCT 257
Db 162839 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACA 162780
QY 258 AGCTGAACAAATTAACAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATT 317
Db 162779 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162720
QY 316 ACGAGAGGGAGAGAGAAATAGAAAAGAAAATGCACATGCTGAATCTGAAGAAGATCAAT 377
Db 162719 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162660
QY 378 GTCTTAAAGATGAAGAGAGAGAGAGAACTCCATGGCTAAAGTCTGGTAAAGAGATGAAA 437
Db 162659 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162600
QY 438 AAGAAACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493
Db 162599 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162574
```

Search completed: December 4, 2002, 09:36:03
Job time : 6073 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen, Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 06:40:33 : Search time 271 seconds
(without alignments)
14125.927 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttatgtgaaca.....tttcgacttacgaacata: 1700

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370476

Minimum DB seq length: 3
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	76.8	4.5	7657	22	AAS45477 Chemically pretreated
C 2	76.8	4.5	7657	24	ABJ34022 Human immune syste
C 3	76.8	4.5	14006	24	ABJ33958 Human immune syste
C 4	76	4.5	19124	18	AA72882 Plasmodium var-7 g
C 5	76	4.5	19124	21	AA298287 Plasmodium var-7 p
C 6	74.8	4.4	9539	22	AAS45347 Chemically pretrea
C 7	74.8	4.4	9539	24	ABK28180 DNA transcription
C 8	74	4.4	6644	20	AAJ33181 Base sequence of t
C 9	74	4.4	7372	20	AAJ33182 Base sequence of t

10	74	4.4	7797	20	AAJ33180 Cowpox virus bsr f
11	74	4.4	7996	20	AAJ33184 Base sequence of t
C 12	73.4	4.3	113515	24	ABL34174 Human immune syste
C 13	72.6	4.3	16033	24	ABL33404 Human immune syste
C 14	72	4.2	16167	24	ABL70254 Chemically treated
C 15	72	4.2	16167	24	ABL33083 Human immune syste
C 16	72	4.2	16167	24	ABL34529 Human metastasis a
C 17	71.6	4.2	17934	24	ABL33319 Human immune syste
C 18	71.2	4.2	10620	20	AAJ22996 Human IL-1ra BAC c
C 19	71.2	4.2	14690	20	AAJ22303 Human IL-1ra BAC c
C 20	70.6	4.2	34769	22	AAS46775 Tumour suppressor
C 21	70	4.1	5750	22	AAS46708 Tumour suppressor
C 22	70	4.1	5750	24	ABL34008 Human immune syste
C 23	69.6	4.1	556	24	ABQ36896 Oligonucleotide fo
C 24	69.6	4.1	556	24	ABQ36997 Oligonucleotide fo
C 25	69.4	4.0	723	24	ABQ25178 Oligonucleotide fo
C 26	69.4	4.0	723	24	ABQ25179 Oligonucleotide fo
C 27	68	4.0	292	24	ABL86605 Human ovarian canc
C 28	68	4.0	6145	24	ABL32972 Human immune syste
C 29	68	4.0	16633	24	ABM79984 Human chemically m
C 30	67.8	4.0	8079	24	ABM79984 Chemically treated
C 31	67.4	4.0	8404	22	AAJ46500 Tumour suppressor
C 32	67.4	4.0	8404	24	ABL33595 Human immune syste
C 33	67.2	4.0	556	23	ABV40063 Human prostate exp
C 34	67.2	4.0	556	23	ABV40163 Human prostate exp
C 35	67.2	4.0	556	23	ABV42105 Human prostate exp
C 36	67.2	4.0	556	23	ABV43601 Human prostate exp
C 37	67	3.9	6668	24	ABL33697 Human immune syste
C 38	66.8	3.9	174424	24	ABJ68122 Ovary cancer relat
C 39	66.6	3.9	6485	22	AAS46559 Tumour suppressor
C 40	66.6	3.9	6485	24	ABL33808 Human immune syste
C 41	66.4	3.9	299	24	ABL86618 Human ovarian canc
C 42	66.4	3.9	563	23	ABV50724 Human prostate exp
C 43	66.4	3.9	6338	24	ABL33126 Human immune syste
C 44	66.4	3.9	16200	24	AAD36438 Mouse L66 genomic
C 45	66.2	3.9	307	24	ABL86438 Human ovarian canc

ALIGNMENTS

RESULT :

10 AAS45477 standard; DNA; 7657 BP.

XX AAS45477:

AC AAS45477:

XX 18-DEC-2001 (first entry)

XX Chemically pretreated complementary DNA associated with cell cycle #91.
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
XX human immunodeficiency virus; neurodegenerative disorder; solid tumour;
XX graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
XX arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
XX immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
XX PCR primer.

XX Homo sapiens.

XX WC200168911-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02945.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019123.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC*) EPiGENOMICS AG.

XX	Olex A. Piepenbrock C. Berlin X;
PI	WPI: 2001-602751/58.
XX	
DR	Designing primers and probes for analysing diseases associated with
XX	cytosine methylation state e.g. arthritis, cancer, aging,
PT	arteriosclerosis comprising fragments of chemically modified genes
PT	associated with cell cycle.
XX	
PS	Claim 1; SEQ ID No 162; 28pp; English.
XX	
CC	Sequences AAS45236-AAS45520 represent chemically pretreated genomic DNA
CC	molecules associated with the cell cycle and specific PCR primers of the
CC	invention. The sequences are useful for detecting the methylation state
CC	of all CpG dinucleotides in a sequence and therefore for analysing
CC	associated diseases. By analysing cytosine methylations in the pretreated
CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC	of existing diseases or the predisposition to specific diseases can be
CC	ascertained. The parameters may be compared to another set of genetic
CC	and/or epigenetic parameters, the differences serving as basis for
CC	diagnosis and/or prognosis events which are disadvantageous to patients.
CC	The sequences of the invention are useful for the diagnosis and therapy
CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC	aging, 9-cmethyl disease, Lewy body disease, arthritis,
CC	arteriosclerosis, solid tumours and cancers.
XX	
SQ	Sequence 7657 BP: 2028 A; 82 C; 1628 G; 3919 T; 3 other;
Query Match 4.53; Score 76.8; DB 22; Length 7657;	
Best Local Similarity 48.64; Pred. No. 6e-06;	
Matches 210; Conservative 0; Mismatches 222; Indels 0; Gaps 0;	
QY	65 ACAATAATTAATATCACTGTTAATAACATGTTTGTGTTCTTAAATACAGCATTTAA 134
DB	
DB	6067 AAAAAATTTAAITACCAACATCAAAATTAATTATACCTATATAATATAAAATCAATT 6008
QY	125 ATTCAGACATTTGTTTTRAAATCAAAATCTAATCTTTATATCAACAGCATTCACGGAA 184
DB	
DB	6007 ACTCCAAAAAACAACAATATCCCGAACATTAAATATATCAACCAAAAAAACGTA 5948
QY	185 ATTCAGGTAAAAAGGAAATTAAGATATGAGATAGAGAGATTTCTATGGAAAAAGAA 244
DB	
DB	5947 ACGCAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 5828
QY	245 AGAGAGAACATGTAGTGAACAAAAATAAGAGATATGATATATATTTATGAGAGTGG 304
DB	
DB	5887 AAAAAAACAACAACAAAAAATAAAATATATATAAATAAAAAAATAAAAAAATAAAA 5828
QY	305 TGAAGATTATTAGAGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTG 364
DB	
DB	5827 AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 5763
QY	365 ACGAATCAATTGTTTAAAGATCAAGAGAGAAAGACATCTCCATGGCTTAAGTCTGGT 424
DB	
DB	5767 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 5708
QY	425 AAAGAAGATTGAATAAGAAACAAAGAAGGAGAGACAAAGAGAAAGGCTTAATATAGCTAA 484
DB	
DB	5707 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 5648
QY	485 CTATGCCAAA 496
DB	
DB	5647 ACTTACTAATA 5636
RESULT 2	
ABL34022/C	
ID	ABL34022 standard; DNA; 7657 BP.
XX	ABL34022;
AC	
XX	
DT	26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 1995.

Human: immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antianemic; cytostatic; nootropic; neuroprotective; anti-HIV; anticonvulsant; ophthalmological; antirheumatic; antiarthritic; antidiabetic; antipsoriatic; antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene: ds.

Homo sapiens.

WO200200928-A2.

03-JAN-2002.

02-JUL-2002; 2001WO-EP07537.

30-JUN-2000; 2000DB-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI: 2002-130909/17.

Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation -

C claim 1; SEQ ID NO 1995; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

Sequence 7657 BP; 2028 A; 82 C; 1628 G; 3919 T; 0 other:

Query Match 4.5%; Score 76.8; DB 24; Length 7657;
Best local similarity 48.6%; Pred. No. 6c-06;
Matches 210; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

65 ACAATATTAAATATCATCAGTGTTAATAACAATGTTCCTTAAATACATGCATTTTAA 124
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
6067 AAAAAATTAATTCACACATCAAAAATTAATTACCATTATAAAT/AAAAAAAAATCAAT 6088
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
125 AATCAGACATTCGTTTTAAATCAAATCTAATCTTATCACACGCACATTCACGAA 184
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
6007 AACCCAAAAAACAACAATATCCCGAACATTAATATATAT/CACCCAAAAAACAAGTA 5948
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
125 AATTCAGGTAAACAGAGAAATTAAGNATGAGAGATAGAGAGATTCATGGAAAAAGAA 244
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
5947 AACGAATAIAAAAAAATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 5868
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
245 AGAGAGAACATGTAGGTGAACAAATAAAGAGATATGATGATATATTTATGAGAGGTG 304
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
5887 AAAAANAANAACAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 5828
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
305 TCAAGATTATTTTAGGAGAGGAGAGAANAATAGAAAAAGAAAATGCATGTGATCTG 364
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
5827 AAAAATAAANAANAANAANAANAANAANAACAANAANAANAANAANAANAANAANA 5768
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
365 AAGAGATGAATTCGTTGAAGATGAAGAGAGAAAGAGAACTCCATGCGCTAAAGTCTCGT 424
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
5767 AAAACCAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 5708
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 2	
ABL34022/	
ID	ABL3
XX	
AC	ABL3
XX	
DT	26-N

PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PT binding proteins

XX PS Claim 4: Page 56-61: 96pp; English.

XX This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
CC var family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and
CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the
CC invention. The compositions are for the treatment and prevention of
CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
CC family of genes having homology with conserved regions of DABP and SABP.
CC The compositions are used for the treatment and prevention of malaria.
CC They are also used in the preparation of vaccines for inducing a
CC protective immune response in a mammal to Plasmodium merozoites
CC (especially Plasmodium falciparum or Plasmodium vivax).

XX SX Sequence 19124 BP; 7624 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 18; Length 19124;
Best Local Similarity 50.0%; Pred. No. 1e-05;
Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;
QY 52 ATTGGATAGCAACACAAATTAATAATATCATGCTTAAATACATGTTTCTCTAAAT 111
DB 15440 ATTATTAANAAGAAAAGAGAAATGAAATATTAANAANAATATTAANAANAANA 15499
QY 112 ACATGCAATTTAAATACAGACATTTGTTTAAATCAAAATCTATATATCATCAAC 171
DB 15500 AAAAAAAG 15559
QY 172 GACATTCAGCGGAAATTCAGGTAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 231
DB 15560 AATTTGTAGATTAANAATGAAAAGAGATTTCAANAANAANAATTAANAANAANA 15616
QY 232 TATGAG 291
DB 15617 TATATAAANAANAATGATTATAAANAANAATTAANAANAANAATTAANAANAANA 15676
QY 292 TATGAG 351
DB 15677 AAAAAAAG 15736
QY 352 CATGGTGAATCTGAAGAGATGAATCTGTTAAAGAGAGAGAGAGAGAGAGAGAGAG 411
DB 15737 AATATATATCATTAANAANAANAATTAANAANAATTAANAANAATTAANAANAANA 15796
QY 412 GCTAAAGTCTGTAAGAGATGAATCTGTTAAAGAGAGAGAGAGAGAGAGAGAGAG 471
DB 15797 AATATAAANAANAATTAATTAANAANAANAATTAANAANAATTAANAANAANA 15856
QY 472 TAAATAGACTACTA 487
DB 15857 AAAAAAATAAATAA 15872

RESULT 5

AZ98287
ID AZ98287 standard; DNA; 19124 BP.

XX AC AZ98287;

XX AC AZ98287;

XX 13-JUN-2000 (first entry)

XX DT

DE Plasmodium var-7 polypeptide encoding DNA.

XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
XW DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
XW protozoan; var-7; ds.

XX Plasmodium sp.

XX CS59993827-A.

XX 30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.

XX 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Sim KL, Chitnis C, Peterson DS, Su X, Wellem TE, Miller LH;

XX WPI: 2000-194198/17.

XX P-PSDB: AAY77604.

XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -

XX Disclosure; Columns 91-108; 93pp; English.

XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunological studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-7 polypeptide encoding
CC DNA.

XX SX Sequence 19124 BP; 7624 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 21; Length 19124;
Best Local Similarity 50.0%; Pred. No. 1e-05;
Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;

QY 52 ATTGGATAGCAACAAATTAATAATATCATGCTTAAATACATGTTTCTCTAAAT 111
DB 15440 ATTATTAANAAGAAAAGAGAAATGAAATATTAANAANAATTAANAANAANA 15499
QY 112 ACATGCAATTTAAATACAGACATTTGTTTAAATCAAAATCTATATATCATCAAC 171
DB 15500 AAAAAAAG 15559
QY 172 GACATTCAGCGGAAATTCAGGTAAAAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 231
DB 15560 AATTTGTAGATTAANAATGAAAAGAGATTTCAANAANAANAATTAANAANAANA 15616
QY 232 TATGAG 291
DB 15617 TATATAAANAANAATGATTATAAANAANAATTAANAANAANAATTAANAANAANA 15676
QY 292 TATGAG 351
DB 15677 AAAAAAAG 15736
QY 352 CATGGTGAATCTGAAGAGATGAATCTGTTAAAGAGAGAGAGAGAGAGAGAGAGAG 411
DB 15737 AATATATATCATTAANAANAANAATTAANAANAATTAANAANAATTAANAANAANA 15796
QY 412 GCTAAAGTCTGTAAGAGATGAATCTGTTAAAGAGAGAGAGAGAGAGAGAGAGAG 471

15797 AAATAAAAAATTTATTAATAGAAAAAATAATATTAATTAATTAATAA 15856
 472 TAAATAGACTAATA 487
 15857 AAAAAATAATAATA 15872

RESULT 5
 AAS45347/c
 AAS45347 standard; DNA: 9539 BP.

AAS45347:

18-DEC-2001 (first entry)

Chemically pretreated complementary DNA associated with cell cycle #26.

Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;
 immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 PCR primer.

Homo sapiens.

W0200168911-A2.

20-SEP-2001.

15-MAR-2001; 2001WO-EP02945.

15-MAR-2000; 2000DE-1013847.

06-APR-2000; 2000DE-1019038.

07-APR-2000; 2000DE-1019173.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;
 WPI; 2001-602751/68.
 Designing primers and probes for analysing diseases associated with
 cytosine methylation state e.g. arthritis, cancer, aging,
 arteriosclerosis comprising fragments of chemically modified genes
 associated with cell cycle.

Claim 1; SEQ ID No 52; 28bp; English.

Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 molecules associated with the cell cycle and specific PCR primers of the
 invention. The sequences are useful for detecting the methylation state
 of all CpG dinucleotides in a sequence and therefore for analysing
 associated diseases. By analysing cytosine methylations in the pretreated
 DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 of existing diseases or the predisposition to specific diseases can be
 ascertained. The parameters may be compared to another set of genetic
 and/or epigenetic parameters, the differences serving as basis for
 diagnosis and/or prognosis events which are disadvantageous to patients.
 The sequences of the invention are useful for the diagnosis and therapy
 of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 aging, glomerular disease, Lewy body disease, arthritis,
 arteriosclerosis, solid tumours and cancers.

Sequence 9539 BP; 2078 A; 121 C; 228 G; 5059 T; 0 other;

Query Match

Best Local Similarity 4.4%; Score 74.8; DB 22; Length 9539;

Matches 208; Conservative C; Mismatches 222; Indels 0; Gaps 0;

QY 67 AAATATTATATACATGTTAATAACATGTTTGTCTTAAATACATGCTTTTAAAA 126
 DB 589 AAAAAAATAATTAACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 530
 QY 127 TCAGACATTTGTTTAAATCAATCTTAATCTTTATATCACAACGACATTGACGAAAA 186
 DB 529 AAAAAAATAATCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 470
 QY 187 TTCAGTAAAGAGAAAAATAAGCAATGAGAGATAGAGATTTCTTATGGAAGAAAGAG 246
 DB 469 TAAAAAATAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 410
 QY 247 AGACACATGTCAGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAGGTG 306
 DB 409 AAAAAAATAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 350
 QY 307 AACATTTATTTAGCAGAGGAGAGAGAAATAGAAAAAGAAAAATGACAIGTGAATCTGAA 366
 DB 349 AAAAAAATAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 290
 QY 367 GAAGATGAATTTGTTTAAATCAATCTTAATCTTTATATCACAACGACATTGACGAAAA 426
 DB 289 AAAAAAATAATCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 240
 QY 427 AGAAGATGAAAAAGAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 486
 DB 229 AAAAAAATAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 170
 QY 487 ATTGCAAAAA 496
 DB 169 AAAAAAATA 160

RESULT 7

ARK28180/c

ID ARK28180 standard; DNA: 9539 BP.

XX ARK28180;

XX 23-APR-2002 (first entry)

DNA transcription associated complementary genomic DNA #27.

DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 immunological disorder; Werner syndrome; developmental disorder;
 psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
 polyglutamine disorder; solid tumour.

OS Unidentified.

PN W0200192565-A2.

PJ 04-DEC-2001.

XX 06-APR-2001; 2001WO-EP03973.

XX 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-090046/12.

New nucleic acids or oligomers, useful for diagnosing or treating diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumours or cancer

Claim 1: SEQ ID No 54; 32bp; English.

The invention relates to a nucleic acid, which comprises a segment of the chemically pretreated DNA of genes associated with DNA transcription from one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for diagnosing or treating diseases associated with DNA transcription (particularly with the methylation status), e.g. adrenocortical deaminase deficiency, viral infection, retroviral infection, Sezary syndrome, haematological disorders, immunological disorders, Werner syndrome, tuberculosis, developmental disorders, psoriasis, Reiter's syndrome, neurological disorders, neurodegenerative disorders, Waardenburg syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial infarction, hypertension, angiogenesis, erythropoiesis, congenital heart disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours or cancer. Sequences ABX28127-ABX28472 represent DNA transcription associated genomic DNA molecules of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from the European Patent Office.

Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 other;

Query Match	4.4%	Score	74.8	DB	24	Length	9539
Best Local Similarity	48.4%	Prod. No.	1.6e-05				
Matches	208	Conservative	0	Mismatches	222	Indels	0
Gaps							
57	AAATATTAAATATATCAGTGTAAATACATGTTTGTGTCCTTAAATATACATGCATCTTTAAAA	126					
589	AAAAAAAAAAATAAAAACAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	530					
127	TCAGACATTTGTTTAAATCAAACTAAATCTCTTATTCACACGACATTTGACGGAAAA	196					
529	AAAAAAAAAAATCAAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	470					
187	TTCCAGGTAAAAAGAGAAAAATAACGAATGACAGATAGAGAGATCTTCTGTAAGAAACGAAAC	246					
469	TAAAAAANAATAAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	410					
247	AGAGACATGTAGGTGAACAAAATAAGAGAGATATGATGATATATTTATGAGAGGTGGT	306					
409	AAAAAAAAACGAAAAAANAAGAAAAAANAANAANAANAANAANAANAANAANAANAANAANA	350					
307	AAGATCTATTTTAGGAGAGGAGAGAGAAATAGAAAAGAAAATGCATGGTGAATCTGAA	368					
349	AAAAAAAAAAAAATAAAAACAAATAAAAATAAAAAANAAGCAAAAAANAANAANAANAANA	290					
367	GAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCCTGTA	426					
289	AAAAACAAACGACAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	230					
427	AGAGATGAAAAGAAACAAANGAGGAGAGAAAGAGAAAGGCTAAATATGACCTAACT	486					
229	ATAAACGAAAAAATACAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	170					

RESULT 8
AX33181
O AAX33181 standard; DNA: 6644 BP.

AC	AAX33181;
XX	
CC	25-JUN-1999 (first entry)
CC	
DE	Base sequence of the plasmid pRx-ires-bsr.
XX	
XX	Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW	crma; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW	autoimmune disease; graft rejection reaction; inflammation;
KW	inflammatory disease; ss.
XX	
XX	Synthetic.
OS	Cowpox virus.
OS	
XX	WO99:3073-A2.
PN	
XX	
XX	18-MAR-1999.
PD	
XX	
XX	07-SEP-1998; 98WO-JP04010.
PF	
XX	
XX	08-SEP-1997; 97JP-0259235.
PR	
XX	
XX	(RPRG-) RPR GENCELL ASIA PACIFIC INC.
PA	
XX	
XX	Hamada H;
P1	
XX	
DR	WPI: 1999-243728/20.
XX	
XX	New apoptosis-resistant virus-sensitive cell
PT	
XX	
XX	Example 1: Page 36-41; 5lpp: English.
PS	
XX	
CC	The present invention describes an apoptosis-resistant virus-sensitive
CC	cell line into which an apoptosis resistance gene has been introduced.
CC	The recombinant viruses generated are capable of expressing apoptosis-
CC	associated genes. These can then be used in a variety of diseases for
CC	which the induction of apoptosis by gene transfer, or where the
CC	inhibition of harmful apoptosis, is therapeutic. The recombinant virus
CC	are useful as vectors for gene therapy which can be applied to cancer
CC	therapy for destroying cancer cells selectively, the treatment of
CC	autoimmune diseases and graft rejection reaction, and apoptosis induc-
CC	therapy for inflammatory cells in inflammatory diseases. Prior arts ha-
CC	encountered the problem where if an adenovirus vector capable of
CC	expressing an apoptosis-associated gene is introduced into anima. cell
CC	the cells producing the virus will be destroyed because the period of
CC	time required to induce cell death by apoptosis is shorter than that
CC	required to replicate and produce the virus, resulting in failure to
CC	obtain a recombinant virus having the integrated apoptosis-associated
CC	gene. In this invention an apoptosis-resistant 293 cell line (having a
CC	apoptosis-resistant gene introduced) is established and overcomes the
CC	problem. The present sequence represents the base sequence of the
CC	plasmid pRx-ires-bsr, which contains the cowpox virus bsr gene, and
CC	is used in an example from the present invention.

Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;

```
Query Match      4.4%; Score 74; DB 20; Length 6544;
Best Loca: Similarity 48.4%; Pred. No. 2.1e-05;
Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
```

Qy	107	AAATACATGCATTTTAAANTCAGACACATTGTTTAAANTCAATCTAATCTCTATAIC	166
Db	4018	AA	4077
Qy	167	ACAACACACATIGACGGAAATTCAGGTAAAAGACAAATATAAGATCAGAGATAGAGAG	226
Db	4078	AA	4137
Qy	227	ATTTCATGTGGAAGAGAGACACATGTAGGTGAACAAATAAAGAGATATGATGAT	286
Db	4138	AA	4197
Qy	287	ATATTTTATGAGAGGTGGTGAAGATTATTTTAGGAGAGGGAGAGAAATAGAAAAAGAA	346

DR WPI: 1999-243728/20.

XX New apoptosis-resistant virus-sensitive cell

XX Example 1: Page 34-38; 5:pp; English.

XX The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the cowpox virus bsr gene which is used in an example from the present invention.

XX Sequence 7797 BP: 2542 A; 1760 C; 1656 G; 1839 T; 0 other:

XX Query Match 4.4%; Score 74; DB 20; Length 7797;

XX Best Local Similarity 48.4%; Pred. No. 2.2e-05;

XX Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 107 AAATACATGCTTTAAATCAGACATTTGTTTAAATCAATCTAATCTATATC 166

DB 5171 AA 5230

QY 167 ACAACGACATTTGCGGAAAAATTCAGGTAAAAAGAGAAATAAAGATGAGAGAG 226

DB 5231 AA 5290

QY 227 ATTTCTATGGAAGAGAGAGAGATGAGTGAACAAATAAAGAGATATGAT 286

DB 5291 AA 5350

QY 287 ATATTATGAGAGTGTGGAAGATTTTAAAGATGAGAGAGAGAGAGAG 346

DB 5351 AA 5410

QY 347 AATGACATGTTGAATCTGAAGAGAGATGATTTGTTTAAAGATGAGAGAGAGAG 406

DB 5410 AA 5470

QY 407 CCATGGCTAAGTCTGTAAGAGAGATGAAAGAGAGAGAGAGAGAGAGAG 466

DB 5471 AA 5530

QY 467 AAGGCTAAATAGACTAATCTGCGGAAAAATTTCTGTAGCGGAGAGAGAG 526

DB 5531 AA 5590

QY 527 AAGGCT 532

DB 5591 AAGGCT 5596

RESULT 11

ID AAX33184

AC AAX33184 standard; DNA: 7996 BP.

XX AAX33184;

XX 25-JUN-1999 (first entry)

XX Base sequence of the plasmid pR-Bcl 2-i-hcd 25.

XX Cowpox virus: bsr: viral vector; expression; apoptosis; resistance;

XX crna: bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;

XX autoimmune disease; graft rejection reaction; inflammation;

XX inflammatory disease; ss.

XX Synthetic.

XX Homo sapiens.

XX W09913073-A2.

XX 15-MAR-1999.

XX 07-SEP-1998; 98WO-JP04010.

XX 08-SEP-1997; 97JP-0259235.

XX (RPRG-) RPR GENCELL ASTA PACIFIC INC.

XX Hamada H;

XX WPI: 1999-243728/20.

XX New apoptosis-resistant virus-sensitive cell

XX Example 3: Page 46-49; 5pp; English.

XX The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the base sequence of the plasmid pR-Bcl 2-i-hcd 25, which contains the human Bcl-2 gene, and is used in an example from the present invention.

XX Sequence 7996 BP: 2463 A; 2015 C; 1829 G; 1669 T; 0 other:

XX Query Match 4.4%; Score 74; DB 20; Length 7996;

XX Best Local Similarity 48.4%; Pred. No. 2.2e-05;

XX Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 107 AAATACATGCTTTAAATCAGACATTTGTTTAAATCAATCTAATCTATATC 166

DB 5370 AA 5429

QY 167 ACAACGACATTTGCGGAAAAATTCAGGTAAAAAGAGAAATAAAGATGAGAGAG 226

DB 5430 AA 5489

QY 227 ATTTCTATGGAAGAGAGAGAGATGAGTGAACAAATAAAGAGATATGAT 286

DB 5490 AA 5549

QY 287 ATATTATGAGAGTGTGGAAGATTTTAAAGATGAGAGAGAGAGAGAGAG 346

DB 5550 AA 5609


```

ConScore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: December 4, 2002, 06:41:38 : Search time 71 Seconds
        (without alignments)
        9336.418 Million cell updates/sec

Title: US-09-733-685.i
Perfect score: 1700
Sequence: 1 tgggtttttattggataacaa.....tttagacttaccacacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 350425 seqs, 194366369 residues
Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

```

ALIGNMENTS

RESULT :
 US 05-983-965-2109/c
 : Sequence 2109, Application US/09983965
 : Patent No. US20020137160A1
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Wesley C.
 : APPLICANT: Jao, Nengbing
 : APPLICANT: Syatt, John C.
 : APPLICANT: Mathialagan, Nagappan
 : TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 : TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
 : FILE REFERENCE: 37-21(10297)C
 : CURRENT APPLICATION NUMBER: US/09/983,965
 : CURRENT FILING DATE: 2001-10-26
 : PRIOR APPLICATION NUMBER: US 09/465,231
 : PRIOR FILING DATE: 1999-12-15
 : PRIOR APPLICATION NUMBER: US 60/113,678

SUMMARY

Result No.	Score	Query		Length	DB ID	Description
		Match				
C 1	73.8	4.3	529	10	US-09-983-965-2109	Sequence 2109, Ap
C 2	72	4.2	446	10	US-09-960-352-3400	Sequence 3400, Ap
C 3	71.2	4.2	152331	9	US-10-095-407-16	Sequence 16, Apr
C 4	71.2	4.2	176373	9	US-10-095-407-17	Sequence 17, Apr
C 5	68	4.0	292	10	US-09-867-701-9583	Sequence 9583, Ap
C 6	68	4.0	516	10	US-09-960-352-5785	Sequence 5785, Ap
C 7	67.8	4.0	180216	10	US-09-835-2332-6	Sequence 6, Apr-1
C 8	66.8	3.9	174424	10	US-09-967-768A-314	Sequence 314, App
C 9	66.4	3.9	299	10	US-09-867-701-9596	Sequence 9596, Ap
C 10	66.2	3.9	307	10	US-09-867-701-9415	Sequence 9415, Ap
C 11	65.6	3.9	42999	9	US-09-799-462A-17	Sequence 17, Apr
C 12	65.6	3.9	42999	9	US-10-125-767-17	Sequence 17, Apr-1
C 13	65.2	3.8	451	10	US-09-960-352-10262	Sequence 10262, A
C 14	63.8	3.8	174424	10	US-09-967-768A-314	Sequence 314, App
C 15	63.4	3.7	429	10	US-09-960-352-5793	Sequence 5793, App
C 16	63	3.7	299	10	US-09-867-701-9477	Sequence 9477, Ap
C 17	63	3.7	341	10	US-09-960-352-12302	Sequence 12302, A
C 18	62.6	3.7	152331	9	US-10-095-407-16	Sequence 16, Apr-1
C 19	62.4	3.7	1721	9	US-09-938-842A-3578	Sequence 3578, Ap


```

; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-17

Query Match      4.2%; Score 71.2; DB 9; Length 176373;
Best Local Similarity 54.6%; Pred. No. 0.0001;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AAATAAAGATGAGATAGAGATTTCTATGTAAGAAAGAGAGACATGTAGGTG 262
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 144774 AGCAAGAAGAAAGAGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAA 144715

QY 263 ACAAAATAAAGAGATATGATGATATATTTATGAGAGGTGGTGAAGATATT 322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 144714 AGAAAGAAAGAAAGAGAGAGAGGAGGAGGAGAGAGAGAGAGAGAGAG 144555

QY 323 AGGAGAGAGAGAAATAGAAAGAAATGACATGGTGAATCTGAGAGAGATGAT 362
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 144654 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144595

QY 383 AAAGATGAAGAGAGAGAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATG 442
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 144594 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144535

QY 443 ACAAAAGAGGAGAGAGAGAA 462
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 144534 AGCAAGGAGAGAGAGAGAA 144515

RESULT 5
US-09-867-701-9583
; Sequence 9583, Application US/39867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9583
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9583

Query Match      4.0%; Score 68; DB 10; Length 292;
Best Local Similarity 52.9%; Pred. No. 7.7e-05;
Matches 146; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 196 AAGACAGAAATAAAGATGAGATGAGATGAGATTTCTATGGAAAAAGAGAGACAT 255
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10 ACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69

QY 256 GTAGGTGAACAAAATAAAGAGATATGATGATATTTATGAGAGGTGCTGAAGATT 315
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 70 GAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 129

QY 316 TTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 375
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 130 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189

QY 376 TTGTGTTAAAGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 130 AAACGAGCGAGCGAGGGGAGAGCGAGAGCGAGAGAGAGAGAGAGAGAGAGAA 249
QY 436 AAAGACAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGC 471
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 250 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTC 285

RESULT 6
US-09-960-352-5785/c
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 5112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; NAME/KEY: unsure
; LOCATION: (76),(9c)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-L-B3057-018-Q1-K1-G1
US-09-960-352-5785

Query Match      4.0%; Score 68; DB 10; Length 516;
Best Local Similarity 48.2%; Pred. No. 8.9e-05;
Matches 217; Conservative 0; Mismatches 232; Indels 1; Gaps 1;

QY 68 AATATTATATATCATGCTGTAAATACATGTTTCTTCTTAAATACATGCAATTTAA 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 501 AAAAAAATATTAAATA:AAAAAATAAATAAATAAATAAATAAATAAATAAATAA 442

QY 123 CAGACATTTGTTTTAAATCTAATCTCTTATATCAACAGCACATTCAGCGAAAA 187
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 441 ATAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 382

QY 188 TCAGGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 381 TAAAAAATAA:AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 322

QY 248 GAGACATGTAGGTCAACAAAAATAAAGAGATATGATGATATATTTATGAGAGGT 307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 321 ATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 262

QY 358 AGATTATTTTAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 261 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 202

QY 368 AGAGTCAATGTGTTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 201 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 143

QY 428 GAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 142 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 83

QY 488 TTGCCAAATTTCTGTACCGCAAAATACT 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 82 ACCCTCNATACCTTATTCCTTAAAGATT 53

RESULT 7
US-09-835-232-6
; Sequence 6, Application US/09835232

```

Patent No. US20020098489A1
GENERAL INFORMATION:
APPLICANT: Leder, Philip
TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: US/09/835,232
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: US 60/136,811
PRIOR FILING DATE: 2000-04-13
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 180216
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(180216)
OTHER INFORMATION: n = A,T,C or G
US-09-835-232-6

Query Match 4.0%; Score 67.8; DB 10; Length 180216;
Best Local Similarity 49.4%; Pred. No. 0.00046;
Matches 205; Conservative 0; Mismatches 207; Indels 3; Gaps 1;
QY 167 ACAACGACATTGACGGAAAATTCAGTAAAGAGAGAAAATAAGAAATGAGAGATAGAGAG 226
DB 85794 ACAGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 85853
QY 227 ATTTCTATGGAAGAAAGAGAGACAGATGTAGTGAACAAAATAAGAGATATGATGAT 286
DB 85854 AGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 85913
QY 287 ATATTATGAGAGGTGGTGGAGATATTTTAGGAGGGGAGAGAGAAATAGAAAAGAA 346
DB 85914 ACAGATAAGAGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAG 85973
QY 347 AATGACATGGTGAATCTCTAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAGAC 426
DB 85974 AAAGACAGACAGACAGACAGACAGACAGAGAAA---GATACAGAAAGAGACAGACAGAG 56030
QY 407 CCATGGCTAAAGTCTCCTAAGAGACATGAAAAGACAAAGAGAGAGAGAGAGAGAGAG 466
DB 86031 AGAAGAGAGAGACAGACAG 86090
QY 467 AAGGCTAAATAGACTTAATTCGCAAAATTTCTGAGCGGACAAATACTATTTGCTGC 526
DB 86091 AGGAGAGAGACAGACTGAGACTGAGTTGGTCTCCCAATGGGGTTAACAGAAATTTTGGTG 86150
QY 527 AAGGTATTTTGTGATCTTTTGAAGTCAGAAAGTTATTTTACATATACATCTTA 561
DB 86151 GAATTAACACCACTTTTATTTCTTACAGAACTTTCTCTATTTAATTTTCTCTA 86215

RESULT 6
US-09-967-768A-314
Sequence 314, Application US/09/967,768A
Patent No. US20020150677A1
GENERAL INFORMATION:
APPLICANT: Augustus, Meena
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-72
CURRENT APPLICATION NUMBER: US/09/967,768A
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US/60/236,109
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,034
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,111
PRIOR FILING DATE: 2000-09-28

NUMBER OF SEQ ID NOS: 325
SOFTWARE: PatentIn version 3.0
SEQ ID NO 314
LENGTH: 174424
TYPE: DNA
ORGANISM: Homo sapiens
US-09-967-768A-314

Query Match 3.9%; Score 66.8; DB 10; Length 174424;
Best Local Similarity 51.3%; Pred. No. 0.00072;
Matches 155; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 169 AACGACATTGACGGAAAATTCAGGTAAGAGAGAGAAAATAAGAAATGAGAGATAGAGAGAT 228
DB 10275 AAAGAAATAGAAAAG 10334
QY 229 TTCATGCAAAAAGATAT 288
DB 10335 AGGAG 10394
QY 289 ATTTATGAGAGGTGGTGAAGATTTATTTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 348
DB 10395 AAAGAAAG 10454
QY 349 TCACATGCTGAATCTGAAGAGAGATGAATTTGTTTAAGATCAAGAGAGAGAGAGAGACTCC 408
DB 10455 AG 10514
QY 409 ATGGCTAAAGTCTCCTAAGAGAGATGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
DB 10515 AGAAG 10574
QY 469 GG 470
DB 10575 GG 10576

RESULT 9
US-09-867-701-9596
Sequence 9596, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Gores, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.497
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 20912
SOFTWARE: FastSeq for Windows Version: 4.0
SEQ ID NO 9596
LENGTH: 299
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-9596

Query Match 3.9%; Score 66.4; DB 10; Length 299;
Best Local Similarity 52.5%; Pred. No. 0.00016;
Matches 145; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
QY 196 AAAGAGAGAAATAAGAAATGAGAGATAGAGAGATTTCTTATGAAAAGAGAGAGAGAGAG 255
DB 10 ACAGAGAGGAG 69
QY 256 GTAGGTCAACAAAATAAGAGAGATATCATGATAATTTTATGAGAGGTGGTGAAGATTAT 315
DB 70 GAGACCGAGAGACAG 129
QY 316 TTAGGAGAGGAG 375
DB 130 GAGAGACAG 189

QY 376 TTGTTTAAGATGAGAGAGAGAGAACTCCATCGCTAAAGTCTCGTAAGAGAGATCA 435
DB 190 AAACGAGCGAGGAGGAGAGAGCGGAGAGCGGAGAGAAAGAAAGAAAGAAAGAA 249
QY 436 AAAAGAAACAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
DB 250 GAAAGAAAGATC 285

RESULT 10

US-09-867-701-9416
Sequence 9416, Application US/09867701
Patent No. US2002013237A1
GENERAL INFORMATION:
APPLICANT: Agilate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 9416
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-9416

Query Match 3.9%; Score 66.2; DR 10; Length 307;
Best Local Similarity 52.8%; Pred. No. 0.00017;
Matches 143; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 198 AGAGAAATTAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 257
DB 23 ACAGAGCGGAG 82
QY 258 AGGTGACAAATAAAGAGATATGATGATATATTTATGAGAGGTGCTGAAGATTTT 317
DB 83 GAGAGCGGAG 142
QY 318 AGGAGAGGAT 377
DB 143 GAGCAG 202
QY 378 GTGTTTAAGATGAG 437
DB 203 ACAGCGGAGGAG 262
QY 438 AAGAAACAAAG 468
DB 263 AAGAAAG 293

RESULT 11

US-09-799-462A-17/C
Sequence 17, Application US/09799462A
Patent No. US20020160970A1
GENERAL INFORMATION:
APPLICANT: Radiaczky, Gyula
Szalay, Aladar
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McAuliffe
STREET: 4250 Executive Square, 7th Floor
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/799,462A
FILING DATE: 10-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/835,682
FILING DATE: 10-APR-1997
APPLICATION NUMBER: 08/695,191
FILING DATE: 07-AUG-1996
APPLICATION NUMBER: 08/682,080
FILING DATE: 15-JUL-1996
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24601-402G
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858-450-8403
TELEFAX: 858-587-5350
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 42999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: <Unknown>
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-799-462A-17

Query Match 3.9%; Score 65.6; DR 9; Length 42999;
Best Local Similarity 52.5%; Pred. No. 0.00085;
Matches 166; Conservative 0; Mismatches 149; Indels 1; Gaps 1;

QY 166 CACACGACATTCGCGGAAATTCAGGTAAAAAGAGAAATAAAGATGAGAGATAGAGA 225
DB 27215 CAGACAGACAGCGGGGAGAGAAAGACAAACAGAGAGAGAGAGAGAGAGAGAGAGA 27156
QY 226 CATTTCTATGAAAAAGAGAGAGAGACATGTAGTGAACAAATAAAGACATATGATGA 285
DB 27155 GACACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAGAGAGAGAGAG 27096
QY 286 TAATTTTATGAGAGGTGTTGAAGATTTATTTTAGGAGGAGGAGAGAGAGAGAGAGAG 345
DB 27095 GAGAGAGTGAAGAGGAGAGAGAGAGAGATGGAGGAGAGAGAGAGAGAGAGAGAGAGA 27036
QY 346 AATGACATGGTGAATCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAGAGAG 405
DB 27035 AACAGACAGAAAGAGAGAGAGAGAGAGCGCGAGAGAGAGTGAAGAGAGAGAGAGAGAG 26976
QY 406 TCCATGGCTAAAGTCTCGTAAAGA-AGATGAAAAAGAGAGAGAGAGAGAGAGAGAGAG 464
DB 26975 TGGAGGGAG 26916
QY 465 CAAAGGCTAAATAGA 480
DB 26915 GTGAGTGAAG 26900

RESULT 12

US-10-125-767-17/c
Sequence 17, Application US/10125767
Patent No. US20020160410A1

[illegible]

GenCore version 5.1.3
 Copyright (c) 1993 - 2002 CompuGen Ltd.
 CM nucleic - nucleic search, using sw model
 R-un on: December 4, 2002, 07:05:43 : Search time 2349 Seconds
 (without alignments)
 13436.936 Million cell updates/sec

US-09-733-685-3
 Perfect score: 1700
 Sequence: 1 tgggttttttttgcgaacaa.....tttcgacttgcgaacatat 1700

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32305132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 5%

Maximum Match 100%

Listing first 45 summaries

Database :

ESI:*

- 1: em_estba:*
- 2: em_esthur:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estor:*
- 17: gb_oss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pla:*
- 21: em_gss_vit:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230	13.5	230	17	BH866477
2	203.8	12.0	531	17	BH866479
3	201.2	11.8	206	17	BH866474
4	185	10.9	237	17	BH814587
5	175.2	10.3	196	17	BH866500
6	173.6	10.2	267	17	BH854168

C	7	128.4	7.6	387	17	BH749275
C	8	101.6	6.0	1101	17	CNS0021J
C	9	96	5.6	660	17	BH183498
C	10	96	5.6	660	17	CNS070N
C	11	91	5.4	982	17	AQ325799
C	12	89	5.2	529	9	AL514657
C	13	88	5.2	769	17	CNS01V5E
C	14	86.6	5.1	1101	17	CNS008VL
C	15	86.4	5.1	764	17	AZ193158
C	16	85.8	5.0	842	17	CNS010QN
C	17	85.6	5.0	997	17	CNS005TE
C	18	85	5.0	886	17	CNS075DF
C	19	85	5.0	934	17	AZ184244
C	20	85	5.0	1201	17	CNS0167M
C	21	84.8	5.0	1101	17	CNS00Lr2
C	22	84	4.9	647	17	BH038922
C	23	84	4.9	839	14	BQ151187
C	24	83.8	4.9	922	17	AZ548363
C	25	83.4	4.9	636	17	AQ329262
C	26	83.4	4.9	846	17	AZ549808
C	27	82.6	4.9	89	17	BH866492
C	28	82.6	4.9	718	17	AG160666
C	29	82.4	4.8	434	17	CNS06Q23
C	30	82.4	4.8	952	17	AQ897460
C	31	82.2	4.8	580	17	CNS008HN
C	32	82.2	4.8	1101	17	CNS0039G
C	33	82	4.8	888	17	AZ549422
C	34	81.9	4.8	625	17	CNS036A2
C	35	81.8	4.8	915	17	AG044049
C	36	81.4	4.8	831	17	CNS009JY
C	37	81.2	4.8	1031	17	CNS0064G
C	38	81.2	4.8	1034	17	CNS06PLB
C	39	81.2	4.8	1101	17	CNS012TP
C	40	81	4.8	89	17	BH866501
C	41	80.8	4.8	1042	17	CNS0148K
C	42	80.6	4.7	797	10	BE034282
C	43	80.6	4.7	1101	17	CNS0181N
C	44	80.4	4.7	614	17	CNS0152H
C	45	79.8	4.7	987	17	CNS00418

ALIGNMENTS

RESULT 1
 BH866477
 LOCUS
 DEFINITION SALK_101380 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_101380, DNA sequence.
 ACCESSION BH866477
 VERSION BH866477.1
 KEYWORDS GSS, GI:22102375
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 230)
 AUTHORS Alonso,J.M., Leisse,I.J., Barajas,P., Chen,H., Cheek,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
 TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10210 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA.

BASE COUNT		genomic DNA inserted into pHOSt using BstXI linkers"			
ORIGIN		153 a	119 c	80 g	179 t
Query Match		12.0%; Score 203.8; DB 17; Length 531;			
Best Local Similarity		72.3%; Pred. No. 1.7e-24;			
Matches 297; Conservative		0; Mismatches 102; Indels 12; Gaps			
QY	590	GTTCGACGGTAAACGAGAGAGAGGTGGATGAGTTTTCAGATATATTACGAGAGACTACAC	1309		
DB	531	GTACGTACCGTACGCGAGGAGAGAGGTGGATGAGTTTTCAGATATATTACGAGAGACTACAC	472		
QY	1010	GTGCGACACGACGAGTTCGGAAGATTACGCGCGGTGCTGTGAGGAGAGATTACCGTCT	1069		
DB	471	GTGCGACACGAGATTGTCGCGAGAGTTACGCGCGGTGAGCTGAACGTGAGATTACCTTCT	412		
QY	1070	AAGAGAGAAACGAGTCCAGATCTCGGTTGAGAAACTCGTGTGATGTGAACGGGCTT	1159		
DB	411	AAGAGAGAAACGGATTTCAGATCTTCGCTGAGGAGCTCATTTGATACTAAGCAACT	352		
QY	1130	CGAGACGGAGATTCGATCAGATTATCGGTCGGGTACAGGGTTTCGGTTTCGATCTG	1188		
DB	351	CAGATGAGAGATCAGATGGAATAATCGGTCGGGTACAGAAATTCGGTTTGATCTT	292		
QY	1190	AACCTGAACCGGACACGACACGCTAGTTTATCGTTGTAGACITGTAGTCCTTCATCT	1249		
DB	291	AACCTGAACCGGACCGCAACGAGTA-----AAATGTAATCTTCTTCATAA	243		
QY	1250	TTTCCCTTCITACAAATCAATTTT-ITTTTAACATACATCTTTTIGAAAAATCG	1308		
DB	242	CTAAATATTTGCCAGTATCTATTTCAATATACATAAAATTTTATTAGCTGG	183		
QY	1309	TAAAGAGACATATTAAACATGTATCCAAATTTCCAGATCTTCAGTTTAT	1359		
DB	182	TTCTAAAGCGTACTAATAATATATCCAGATTCAGATCTCTCTTTCAT	132		
RESULT 3					
LOCUS	BH866474 206 bp DNA linear GSS 05-AUG-				
DEFINITION	SALK_101377 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_101377, DNA sequence.				
ACCESSION	BH866474				
VERSION	BH866474.1 GI:22102372				
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrin,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,Zimmerman,J. and Ecker,J.R.				
AUTHORS	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome				
TITLE	Published (2001)				
JOURNAL	Contact: Joseph R. Ecker				
COMMENT	Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 1001C N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged. Location/Qualifiers 1..206 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_101377"				
FEATURES					
SOURCE					

```
/clone_lib="Arabidopsis thaliana TUNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TUNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tuna\_protocols.html"
```

```

59 a 31 c 72 g 45 t
      GIGIN
      Query Match 11.8%; Score 201.2; DB 17; Length 206;
      Best Local Similarity 98.5%; Pred. No. 6.9e-24;
      Matches 203; Conservative 0; Mismatches 3; Gaps 0

```

938 TCGACGGAGTTGTTCAGACGGTAAACCGAGGAGAGAGTGGATGAGTTTTCAGAGATATTA 997
 1 TCGACGGACCTTGTATCGGACGGTAAACCGAGGAGAGAGTGGATGAGTTTTCAGAGATATTA 60
 948 CGGAGAGTACAGTTCGGCACAGCAACGGTTGCGAAAGTTAACGGCGGTTTTCCTCAGCGA 2057
 61 CGGAGAGTACAGTTCGGCACAGCAACGGTTGCGAAAGTTAACGGCGGTTTTCCTCAGCGA 220
 1058 GAGTTACCGTCTTAAGACAGCAACCGAGTACAACTCTTGGTTGAGAAACTCTCTTGGAT 2117
 121 GAGTTACCGTCTTAAGACAGCAACCGAGTACAACTCTTGGTTGAGAAACTCTCTTGGAT 180

1118 TGTAACGGCGTTCGAGACGGAGAATT 1143
|||||
18: TGTAACGGCGTTCGAGACGGAGAATT 206

RESULTS	237 bp	DNA	linear	SSS 02 MAY-200
BB14567	Arbidopsis	thaliana	TEN4 insertion	lines Arabidopsis
SALK_066574	Arbidopsis	thaliana	gomic clone	SALK_066574. DNA sequence.

ORGANISM

JOURNAL
OF
HERPETOLOGY
Published (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

```

Class: tDNA tagged..
FEATURES             source
    Location/Qualifiers
        1..237
            /organism="Arabidopsis thaliana"
            /strain="Columbia 0"
            /db_xref="taxon:3702"
            /clone="SALK_056674"
            /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
            /note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more T-DNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can

```

```

BASE COUNT      82 a   39 c   63 g   53 t   1 others
be found at http://signal.salk.edu/dna\_protocols.html

```

QY	733	ATATCGACGGTCTAAGTTCACACACGCGCTATAGAGTTTCATTATAAAATTTAGCAA	792
DB			
DB	1	ATATCGACGGTCTAAGTTCACACGCGCTATAGAGTTTCATTATAAAATTTAGCAA	60
QY	795	AAATAAAATCAGCAAAATAATTTTTCCTTGCTAAGCTTAAACGACGCGGTTAAACATTTCCT	852
DB			
DB	61	AATAAAATCAGCAAAATAATTTTTCCTTGCTAAGCTTAAACGACGCGGTTAAACATTTCCT	120
QY	855	TCTGGCTTAACNTGAACAACCTCTTTGAAGAAAGAGAACCGGTAGACAAGATACGGAA	912
DB			
DB	121	TCTGGTASGNTTGGACCAATCTTTGAAGAAAGAGAACCGGTAGACAAGATACGGAG	180
QY	912	AACTCGACGGTATACAGAGGGAACCGTTCGACGGAAGTTGTTCGGACGGTA	961
DB			
DB	181	AACTCGACGGTATACAGAGGGAACCGTTCGACGGAAGTTGTTCGGACGGTA	230

RESULT 5	LOCUS	DEFINITION	LINEAR	DATE
5886552	SH866500	196 bp	DNA	GSS 05-AUG-2002
	SALK101406	Arabidopsis thaliana	RNA insertion lines	Arabidopsis
		thaliana genomic clone	SALK101406	DNA sequence

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

TITLE: Zammareni, J. and Eckert, R. K.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL: Unpublished (2001)
COMMENT: Contact: Joseph R. Eckert
Saik Institute Genomic Analysis Laboratory (SIGAL)
The Saik Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

```

email: eckert@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1. 196
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_lib="SALK_101406"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
BASE COUNT      50 a      33 c      64 g      49 t
ORIGIN
Query Match      10.3%      Score 175.2      DB 17      Length 196
Best Local Similarity 93.4%      Pred. No. 1.5e-19

```



```

AQ325799/c
LOCUS      AQ325799                982 bp    DNA             Linear      GSS 03-JAN-1999
DEFINITION nrx0021B14r CUG1 Rice BAC Library Oryza sativa genomic clone
ACCESSION  AQ325799
VERSION    AQ325799
KEYWORDS   GSS.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzoideae; Oryza.
            1 (bases 1 to 982)
            Wang, R.A. and Dean, R.A.
            A BAC End Sequencing Framework to Sequence the Rice Genome
            Unpublished (1998)
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAACAAGCATGACCATG
            Class: BAC ends
            High quality sequence start: 4
            High quality sequence stop: 123.
            Location/Qualifiers
                1..982      "Oryza sativa"
                /organism="Oryza sativa"
                /strain="Japonica"
                /cultivar="Nipponbare"
                /db_xref="taxon:4530"
                /clone="nrx0021B14r"
                /clone_lib="CUG1 Rice BAC Library"
                /tissue_type="Leaf"
                /lab_host="E. coli DH10B"
                /note="Vector: pReloBAC11; Site:1: HindIII; Site:2:
                HindIII; Rice is one of two most popular grains in the
                world. Half of the world population especially those
                inhabiting highly populated areas of the humid tropics
                and subtropics, rely on rice as their primary source of
                carbohydrate. Monocotyledonous rice is a diploid plant
                (2n=24) with a haploid genome equivalent of 431 Mbp
                (Arumuganathan and Earle, 1991). The relatively small
                genome of rice, three times larger than that of
                Arabidopsis, makes it suitable for genomic studies. In
                order to facilitate positional cloning, physical mapping
                and genome sequencing of rice, we have constructed a BAC
                library from Oryza sativa, Nipponbare variety. The
                library contains 36,864 clones with an average insert size
                of 128.5 Kb providing 10.9 haploid genome equivalents. The
                deep coverage allows the isolation a particular sequence
                with a probability of 99.9 %. Two high density filters,
                each containing 18,432 clones (doubly spotted), represent
                the whole library for colony screening."
                141 a      69 c      43 g      674 t      55 others
BASE COUNT
ORIGIN
Query Match      5.4%; Score 91; DB 17; Length 982;
Best Local Similarity 46.9%; Pred. No. 7.7e-06;
Matches 229; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

QY 9 TATTGATACATGACAATATTTATTTATTCATGACGTTTATTTGGATACATGACACA 58
DB 903 TATGGAAATATAGAAAAGAAATTAATTAATAAATAATTTAATTAATTTAATAAT 844
QY 69 ATATTATATATACAGTCTTAATAACATGTTTCTTCTTAACATACAGCATTTAAATC 128
DB 843 TGTATATATATATATAA-TATATATATTTATATAAA-ATATATATTAATTAATTAAT 784
QY 129 AGACATTTGTTTAAATCAATCTTAATCTTTATATATACACAGCATTCAGCGAAATC 188

```

```

DB 783 AATAATNIATTCGAATAATAATAATAATAATTTTAAATATATATATAATAATAA 724
QY 189 CAGGTAAAGAGAGAAATAAAGATGAGAGTAGAGAGATTTCTTAAGAAAAAGAGAG 248
DB 723 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 664
QY 249 AGAATCATGTAGGTGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGGTG 308
DB 663 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 604
QY 309 GATTATTTAGCAGGAGGAGAGAAATAGAAAAAGAAATGACATGTTGAACTGAAGA 368
DB 603 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 544
QY 369 AGATGAATTTGTTTAAATCATGACAGAGAGAGAAAGACAACTCCATGGCTAAAG 428
DB 643 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 464
QY 429 AAGATGAAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
DB 483 AAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 424
QY 483 TGGCAAAA 496
DB 423 AAAAAAAA 416

```

RESULT 12
 A:514657
 LOCUS A514657 LT_NFL006.PL2 Homo sapiens cDNA clone CLOB0022E09.3
 DEFINITION prime, mRNA sequence.
 ACCESSION A514657
 VERSION A514657.1 GI:12778151
 KEYWORDS EST.
 SOURCE Human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 529)
 Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CLOB0022E09"
 /clone_lib="LFL_NFL006.PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 364 a 20 c 9 g 89 t 47 others
BASE COUNT
ORIGIN
Query Match 5.2%; Score 89; DB 9; Length 529;
Best Local Similarity 43.9%; Pred. No. 2.1e-05;
Matches 206; Conservative 35; Mismatches 228; Indels 0; Gaps 0;

QY 28 TATTATTATTTCAIGAGTGTATTTTTCGACGACGACAAATATATATATACAGTGT 87

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoR digestion of Drosophila DNA and provided by the RBCP from the isogenic strain Y2; cn bw sp. the same strain used for the BCGP's p1 and EST libraries. A more detailed description of the library and how to order individual PAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source 1. 1101
 /organism="Drosophila melanogaster;"
 /db_xref="taxon:7227"
 /clone="BACR29B23"
 /clone_lib="RPCI-98"
 /note="end : 77"
 BASE COUNT 419 a 91 c 60 g 299 t 232 others
 ORIGIN

Query Match 5.1%; Score 86.6; DB 17; Length 1101;
 Best Local Similarity 35.8%; Prod. No. 4c-05;
 Matches 215; Conservative 96; Mismatches 270; Indels 4; Gaps 2;

```

100 5 TTTTATGGATACATGACAAATATTATTATTCATGAGCTTTTATTCATGAGCAATG 64
101 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 b 487 TTATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 546
103 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 65 ACAATATTATATATCATGTTTATATACATGTTTCTCTTAATAATACATGATTTAA 124
105 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 Db 547 ATAACTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 606
107 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
108 125 AATCAGACATTTGTTTAAATCAATCTATCTCTTATATCAATGACATGACGCGAA 184
109 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 Db 607 ATAAAAAATATTTTATTAATAAATTTTATTAATAAATTAATAAATAAATAA 666
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
112 185 AATTCAGTAAAGAGAGAAATAAAGATGAGAGATAGAGAGATTTCTATGAGAAACAA 244
113 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 Db 667 MWMTAATTAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 726
115 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 245 AGAGAGACATGTAGTGAACAAATAAAGAGATGATGATATTTATGAGAGCGTGG 304
117 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 Db 727 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 786
119 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 305 TGAAGATTATTTAGGAGCGGAGAGAGAAATAGAAAAGAAAATGACATGGTGAATCG 364
121 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 Db 787 AWAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 846
123 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 365 AAGAGATGAATGTTGTTAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
125 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 Db 847 WAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 903
127 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 425 AAAGAAGATGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 484
129 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 Db 904 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 963
131 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 485 CTATTCGCAAAATTTCTGACGCAAAATCTACTATTTGCTCAAGGTATTTTGTGTA 544
133 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 Db 964 TTWATTWAAATTAATTTWTTTATTAATTAATTAATTAATTAATTAATTAATTA 1022
135 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 545 GTTTCGAGTCAAAAGTTATTCTTACATATACCTGTAATAAATAAATAAATAA 589
137 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
138 Db 1023 TTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1067
139 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 15

LOCUS AZ193158/c 764 bp DNA linear GSS 30-AUG-2000
 DEFINITION SP_1022_B2_E10_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate-1022 Col-20 Row-J, DNA sequence.
 ACCESSION AZ193158
 VERSION AZ193158.1 GI:8736337
 KEYWORDS GSS.

SOURCE

ORGANISM Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinidea; Euechinoida; Echinacea; Echinoida;
 Strongylocentrotidae; Strongylocentrotus.

REFERENCE

1 (bases 1 to 764)
 Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,I.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Eftensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.
 A sea urchin genome project: Sequence scan, virtual map, and additional resources
 Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)

COMMENT

20402566
 Contact: Cameron, RA, Davidson, EH, Hood, L
 Division of Biology 156-29
 California Institute of Technology
 Pasadena California 91125, USA
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 1022 row: J column: 20
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 764.
 Location/Qualifiers

FEATURES

source : 764
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone="Plate-1022 Col-20 Row-J"
 /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
 /note="organ: sperm; Vector: BAC3.5; BAC Clones in E-Coli DH108"
 BASE COUNT 173 a 56 c 48 g 469 t 18 others
 ORIGIN

Query Match 5.1%; Score 86.4; DB 17; Length 764;
 Best Local Similarity 47.6%; Prod. No. 5e-05;
 Matches 234; Conservative 0; Mismatches 258; Indels 0; Gaps 0;

```

100 5 TTTTATGGATACATGACAAATATTATTATTCATGAGCTTTTATTCATGAGCAATG 64
101 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 Db 728 TATTTAATATAGATATAATAGATTTATTTATTTAGTAATAATANNITGAAGAAT 669
103 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 65 ACAATATTATATATCATGTTTATTAACATGTTTGTCTTAAATAACATGATTTAA 124
105 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 Db 669 AATAATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 609
107 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
108 125 AATCAGACATTTGTTTAAATCAATCTATCTCTTATATCAACAGCATTTGACGCGAA 184
109 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 Db 608 TTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 549
111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
112 185 AATTCAGTAAAGAGAGAAATAAAGATGAGAGATAGAGAGATTTCTATGGAAGAAG 244
113 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 Db 548 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 489
115 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 245 AGAGAGCAATGTAGTGAACAAATAAAGAGATATGATGATATTTTATGAGAGCTGG 304
117 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 Db 488 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 429
119 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 305 TGAAGATTATTTAGGAGCGGAGAGAGAAATAGAAAAAGAAAAATGACATGTAATCTG 364
121 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 Db 428 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 369
123 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
124 365 AAGNAGATGAATGTTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
125 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 Db 368 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 309
127 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 425 AAGAGCATCAAAAGAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
129 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 Db 308 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 249
131 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY 485 CTATTGCCAAA 496
| III:
bb 248 AAAAAAAAAA 257

Search completed: December 4, 2002, 09:21:06
JOB time : 2072 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Computer Ltd.

OM protein - protein search, using sw mode!

Run on: December 4, 2002, 07:18:40 : Search time 71 seconds
(without alignments)
228.966 Million cell updates/sec

Title: US-09-733-685-2
Perfect score: 623
Sequence: 1 MNSLKKEERVEHNGKSDG.....QGLCLD..MCKPEPSVSLSL 122

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 13325620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_010021

Rank	Score	Query Match	Length	ID	Description
1	80.5	12.9	265	ABR00653	Chlamydia pneumoniae
2	80.5	12.9	866	22	Amino acid sequence
3	80.5	12.9	968	21	Polycystic kidney
4	80.5	12.9	968	22	Amino acid sequence
5	80	12.8	753	20	N. meningitidis st
6	79.5	12.8	2001	22	Arabidopsis thalia
7	79.5	12.8	2001	22	Arabidopsis thalia
8	78	12.5	245	21	Arabidopsis thalia
9	78	12.5	250	21	Arabidopsis thalia
10	78	12.5	252	21	Arabidopsis thalia

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.5	12.9	265	ABR00653	Chlamydia pneumoniae
2	80.5	12.9	866	22	Amino acid sequence
3	80.5	12.9	968	21	Polycystic kidney
4	80.5	12.9	968	22	Amino acid sequence
5	80	12.8	753	20	N. meningitidis st
6	79.5	12.8	2001	22	Arabidopsis thalia
7	79.5	12.8	2001	22	Arabidopsis thalia
8	78	12.5	245	21	Arabidopsis thalia
9	78	12.5	250	21	Arabidopsis thalia
10	78	12.5	252	21	Arabidopsis thalia

Rank	Score	Query Match	Length	ID	Description
11	77	12.4	685	22	ABG14755
12	77	12.4	966	23	ABR07819
13	77	12.4	966	23	AAW51861
14	76.5	12.3	230	21	AAW04695
15	76.5	12.3	241	21	AAW04694
16	76.5	12.3	249	21	AAW04693
17	76.5	12.3	278	21	AAW04683
18	76	12.2	605	22	ABG05038
19	75	12.0	1843	22	ABG14961
20	74	11.9	141	21	AAW04650
21	74	11.9	808	23	AAO21528
22	73.5	11.8	254	22	ABG04191
23	73.5	11.8	611	22	ABG11703
24	73.5	11.8	2059	23	ABR25711
25	73	11.7	837	22	ABG00837
26	73	11.7	837	22	ABG12133
27	72.5	11.6	666	23	ABP26300
28	72	11.6	255	21	AAW04629
29	72	11.6	260	21	AAW04628
30	72	11.6	262	21	AAW04627
31	71.5	11.5	167	22	ABR65525
32	71.5	11.5	650	22	ABG04247
33	71.5	11.5	662	20	AAW34751
34	71	11.4	217	19	AAW0748
35	71	11.4	217	22	AAW27779
36	71	11.4	217	22	AAW64577
37	71	11.4	215	22	ABG17196
38	71	11.4	2143	22	ABG01716
39	70.5	11.3	376	21	AAW06245
40	70.5	11.3	386	21	AAW06244
41	70.5	11.3	423	23	ABP28260
42	70.5	11.3	551	22	ABR33244
43	70.5	11.3	556	22	ABG03854
44	70.5	11.3	850	17	AAW91308
45	70.5	11.3	859	15	AAW53268

ALIGNMENTS

RESULT 1
ABR00653
ID: ABR00653 standard; Protein; 265 AA.
AC: ABR00653;
XX
XX
DT 29-JUL-2002 (first entry)
DE Chlamydia pneumoniae cp6745 protein, SEQ ID NO:255.
KW Chlamydia pneumoniae cp6745 protein, immunogen; vaccine; diagnosis;
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029.
XX
XX Chlamydia pneumoniae.
XX
XX W0200202606-A2.
XX
XX 10-CAN-2002.
XX
XX 03-JUL-2001; 2001WO-1801445.
XX
XX 03-JUL-2000; 2000GB-0016363.
XX 11-JUL-2000; 2000GB-0017047.
XX 21-JUL-2000; 2000GB-0017983.
XX 07-AUG-2000; 2000GB-0019368.
XX 18-AUG-2000; 2000GB-0020440.
XX 14-SEP-2000; 2000GB-0022583.
XX 10-NOV-2000; 2000GB-0027549.
XX 22-DEC-2000; 2000GB-0031706.

(CHIR) CHIRON SPA.
 Ratti G, Grandi G;
 WPI: 2002-154726/20.
 N-PSDB: ABL91311.
 Novel Chlamydia pneumoniae protein useful in the manufacture of a
 medicament for treatment or prevention of infection due to Chlamydia.
 preferably Chlamydia pneumoniae, and for diagnostic purposes
 Claim 1: Page 152: 364pp; English.
 Sequences AB90926-AB909715 represent novel proteins from Chlamydia
 pneumoniae (strain CW.029), and AB909184-AB909373 represent DNA encoding
 them. The proteins are predicted to be immunogenic and may therefore be
 useful in vaccine production and for diagnostic purposes. Chlamydia
 pneumoniae is a common cause of respiratory disease in humans, and is
 also involved in the development of cardiovascular diseases such as
 atherosclerosis, coronary artery disease, carotid artery stenosis,
 myocardial infarction, cerebrovascular disease, aortic aneurysm,
 claudication and stroke. The proteins and nucleic acids of the invention
 may be used in vaccines and pharmaceutical compositions for the
 prevention or treatment of chlamydial infections, particularly Chlamydia
 pneumoniae infections. The proteins may also be used in the detection of
 Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
 DNA probe assay or blotting techniques for determining Chlamydia
 pneumoniae gene expression. The present sequence represents a
 specifically claimed Chlamydia pneumoniae protein of the invention.

Sequence 265 AA;

Query Match 12.9%; Score 80.5; DB 23; Length 265;
 Best Local Similarity 27.2%; Pred. No. 1.8;
 Matches 40; Conservative 19; Mismatches 37; Indels 51; Gaps 7;
 12 EDNGKSGNRGK-----PTEVVRITVEE---VDEFFKILR-RVHVRRTVAKYNGG 61
 104 QEDLGRVIGRTPEDPFVDITTEITVGLLPDLSELSTVDALOGVRSRLIYVRSVK----- 159
 62 VAEGELPKKKRRKNSQLGLRNSLD-----CNGVR-----DGEF 94
 160 -----PMIQDLAIWGLRSDAQLNFRVLANGVQNHYPHTKVKLYLAKNLADWDCEI 213
 95 DEINRVGLQGLCLDINCKPEPDSVLS 121
 214 SEERKQKALGJD-----PK:ESLSL 236

RESJLT 2
 AAB68448
 ID AAB68448 standard; protein: 866 AA.
 AAB68448;
 23-JUL-2001 (first entry)

Amino acid sequence of an internal fragment of human PKD2.
 PKD2: ADPKD; autosomal dominant polycystic kidney disease;
 cyst formation.
 Homo sapiens.
 Key Location/Qualifiers
 FT Misc-difference 384
 FT Misc-difference 395 /note= "unspecified residue"
 FT Misc-difference 426 /note= "unspecified residue"
 FT Misc-difference 432 /note= "unspecified residue"
 FT Misc-difference 432 /note= "unspecified residue"

FT Misc-difference 547 /note= "unspecified residue"
 FT Misc-difference 586 /note= "unspecified residue"
 FT Misc-difference 587 /note= "unspecified residue"
 FT Misc-difference 588 /note= "unspecified residue"
 FT Misc-difference 849 /note= "unspecified residue"
 FT Misc-difference 849 /note= "unspecified residue"

US6228591-B1.
 06-MAY-2001.
 30-MAY-1999; 99US-0385752.
 23-MAY-1996; 96US-0651999.
 (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 Semlo S, Kochizuki T;
 WPI: 2001-342670/36.

Detecting a mutation in the PKD2 gene, useful for diagnosing autosomal
 dominant polycystic kidney disease, comprises determining the
 difference between the polynucleotide sample and the reference wild
 type PKD2 gene sequence
 Disclosure: Columns 11-18; 34pp; English.

The present sequence represents a fragment of a human PKD2 protein.
 The specification describes a method for detecting the presence of a
 mutation in the PKD2 gene. The method comprises comparing and
 determining the difference between the polynucleotide sample and the
 reference wild type PKD2 gene sequence, where the differences are
 mutations comprising one or more deletion, insertion, point or
 rearrangement mutations. The method is useful for detecting mutation
 in the PKD2 gene and for diagnosing ADPKD (autosomal dominant polycystic
 kidney disease). The method may also be used to determine whether
 persons in the population at large have ADPKD, for identifying persons
 at risk in developing the disease, e.g. relatives of the person with
 ADPKD, as well as for confirming diagnosis of ADPKD. The method may
 further be used for diagnosing ADPKD before clinical manifestations of
 the disease, i.e. formation of cysts.

Sequence 866 AA;

Query Match 12.9%; Score 80.5; DB 22; Length 866;
 Best Local Similarity 28.9%; Pred. No. 8;
 Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
 6 EERVEEDNGKSGNRGKPSIEVVRITVEEVEFFKILRRV----HVATRTVAKYNGVA 63
 710 EDDDEDGHSRRRGSISSGV-----SYEEFQVLRVRDRMEHSIGSVKIDAVIV 762
 64 EGELPSK-KKRRSQNLGLRNSLQNGV-----RDGEFDEINRVGQ----- 103
 763 KLEMERAKLRREVLG--RLDQ--GVAEDRLGROSEIHRQOMERLVREELERWESDDA 818
 194 -----SLG--LDLNCKPEPS 117
 819 ASQISHSLGIPVGLNGOPRPRS 840
 RESJLT 2
 AAY78946
 ID AAY78946 standard; Protein: 968 AA.
 AAY78946;
 05-JUN-2000 (first entry)

X E Polycystic kidney disease PKD2 amino acid sequence.
 X W Polycystic kidney disease; PKD2; ADPKD; treatment: detection;
 X W Autosomal dominant polycystic kidney disease.
 X S Homo sapiens.
 X S US6031089-A.
 X D 29-FEB-2000.
 X F 23-MAY-1996; 96US-0651999.
 X R 23-MAY-1996; 96US-0651999.
 X R (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 X I Somlo S, Mochizuki T;
 X I WPI: 2000-205227/18.
 X R N-PSDB: AAZ95245.
 X T New PKD2 nucleic acid for diagnosis and treatment of autosomal dominant
 T polycystic kidney disease associated with mutations in the PKD2 gene -
 T Claim 1: Fig 5: 35pp; English.
 X C This sequence represents the human polycystic kidney disease (PKD2)
 C protein sequence. The PKD2 gene is associated with Autosomal dominant
 C Polycystic Kidney Disease (ADPKD). The PKD2 gene and protein can be used
 C for the diagnosis and treatment of ADPKD caused by defects in the PKD2
 C gene. ADPKD can be diagnosed by detecting the presence of a mutated PKD2
 C gene in nucleic acid sequences from an individual. Diagnosis of high risk
 C patients prior to the appearance of clinical manifestations is also
 C possible. Detection is carried out by sequencing, restriction enzyme
 C digestion analysis or hybridization of probes to wild type or mutated
 C PKD2 gene or by detecting expression of the gene product i.e. the mRNA or
 C protein product. ADPKD caused by a mutated PKD2 gene can be treated by
 C delivery and expression of a functional PKD2 gene into a sufficient
 C number of cells of the subject to treat the disease.

XX OS Homo sapiens.
 XX PN US6228591-B1.
 XX PD 08-MAY-2001.
 XX PF 30-AUG-1999; 99US-0385752.
 XX PR 23-MAY-1996; 96US-0651999.
 XX PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 XX PI Somlo S, Mochizuki T;
 XX DR WPI: 2001-342670/35.
 XX DR N-PSDB: AAF85440.
 XX PT Detecting a mutation in the PKD2 gene, useful for diagnosing autosomal
 PT dominant polycystic kidney disease, comprises determining the
 PT difference between the polynucleotide sample and the reference wild
 PT type PKD2 gene sequence -
 XX PS Disclosure: Fig 5A-G: 34pp; English.
 XX CC The present sequence represents a human PKD2 protein. The specification
 CC describes a method for detecting the presence of a mutation in the PKD2
 CC gene. The method comprises comparing and determining the difference
 CC between the polynucleotide sample and the reference wild type PKD2 gene
 CC sequence, where the differences are mutations comprising one or more
 CC deletion, insertion, point or rearrangement mutations. The method is
 CC useful for detecting mutation in the PKD2 gene and for diagnosing ADPKD
 CC (autosomal dominant polycystic kidney disease). The method may also be
 CC used to determine whether persons in the population at large have ADPKD,
 CC for identifying persons at risk in developing the disease, e.g. relatives
 CC of the person with ADPKD, as well as for confirming diagnosis of ADPKD.
 CC The method may further be used for diagnosing ADPKD before clinical
 CC manifestations of the disease, i.e. formation of cysts.
 XX SQ Sequence 968 AA;
 Query Match 12.9%; Score 80.5; DB 22; Length 968;
 Best Local Similarity 28.9%; Pred. No. 9.2;
 Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
 QY 8 EERVEEDNGKSDNGRKPSTEVVTVTVEEVEDEFFKILRRV-----HVAITRVAKYNGVA 63
 DB 813 EEDDEDSGHSRRRGSTISGGV-----SYEEFQVLVRRVDRMERHSIGSVSKIDAVIV 865
 QY 64 EGELEPSK-KRRKRSQNLGRNSLDCNGV-----RDGEFDEINRVGLQ----- 103
 DB 866 KLEIMERAKLRREVLS--RLLD--GVAEDERLGRDSEIHREQMERLVREELERWESDUA 921
 QY 104 -----GLG--LDLNCKPEPDS 117
 DB 922 ASQISHGLGTPVGLNQPRPS 943
 RESULT 5
 AAW93495
 ID AAW93495 standard; Protein; 753 AA.
 XX AC AAW93495;
 XX AC AAW93495;
 XX DT 11-JUN-1999 (first entry)
 XX DE N. meningitidis strain M990 LbpB protein.
 XX DE LbpB; lactoferrin binding protein; vaccine; neisserial disease;
 KW meningitis; diagnosis; treatment.
 XX CS Neisseria meningitidis.
 XX CS

PN W09909176-AL.
 XX 25-FEB-1999.
 XX 10-AUG-1998: 98WC-EP05117.
 XX 05-FEB-1998: 98CB-0002544.
 PR 15-AUG-1997: 97CB-0017423.
 XX (UYUT-) RIJKSUNIV UTRECHT.
 PA (TECH-) TECHNOLOGY FOUND TECHNOLOGIESTICHTING ST.
 XX Pettersson-Fernholm AM, Icmasssen JPM;
 XX WPI: 1999-190165/16.
 DR N-PSDB: AAX23322.
 XX New lactoferrin-binding protein B polynucleotides - obtained from
 PT Neisseria meningitidis. Used to develop products for the diagnosis,
 PT prevention and treatment of neisserial disease, e.g. meningitis
 XX Claim 13: Page 102-105: 116pp; English.
 PS This invention describes novel lactoferrin-binding protein B (Lbpb;
 XX strains of Neisseria meningitidis. The products of this invention can
 CC be used for vaccinating humans against neisserial disease e.g.
 CC meningitis. Antibodies raised against the proteins of the invention
 CC can be used for diagnosing or treating neisserial disease in humans.
 CC The Lbpb polypeptides can also be used for identifying compounds which
 CC inhibit the polypeptides.
 XX Sequence 753 AA:
 SQ
 Query Match 12.8%; Score 80; DB 20; Length 753;
 Best Local Similarity 27.8%; Pred. No. 7.6;
 Matches 25; Conservative 18; Mismatches 33; Indels 14; Gaps 3;
 JY 6 KKEERVEEDNGKSDGNRGPSTVEVPTVTEEEVDFFKILRRVHVATRTVAKVNGVAEG 65
 DB 496 EDEETSEEDNGEDETATEETEEVDEAEDEVEE-----PEEKSPAGCNGG-SGS 545
 QY 66 ELPSKKRKSSQNLGRNSLDCNSVRDGEFD 95
 DB 546 ILPALEASKGRDI----DLFLKGIPTAETD 571
 RESULT 6
 ID AAB20062 standard; Protein: 2001 AA.
 XX AAB20062:
 XX 23-APR-2001 (first entry)
 XX Arabidopsis thaliana silencing gene-uncoded protein.
 XX Gene silencing; silencing gene; MOM.
 XX Arabidopsis thaliana.
 XX Key Location/Qualifiers
 FH Region
 FT /note= "internal repeat region"
 FT Peptide
 FT 362..367
 FT /note= "nuclear localization motif"
 FT Binding-site 450..467
 FT /note= "ATP/GTP-binding motif"
 FT 479..719
 FT /note= "region of homology to Arpase/helicase
 FT family SWI2/SNF proteins"
 FT Misc-difference 705
 FT /note= "Lys in ecotype Columbia"
 FT Peptide 832..838

FT Peptide /note= "nuclear localization motif"
 FT 858..862
 FT /note= "nuclear localization motif"
 FT 995..1015
 FT /note= "predicted membrane-spanning domain"
 FT Misc-difference 1219
 FT /note= "Asp in ecotype Columbia"
 FT 1462..1672
 FT /note= "internal repeat region"
 FT 1848..1894
 FT /note= "internal repeat region"
 FT 1899..1941
 FT /note= "actin-binding domain"
 XX W020010601 A2.
 PN 04-JAN-2001.
 PC 21-JUN-2000: 2000WC-EF05761.
 FF 21-JUN-1999: 99CB-0014623.
 PR (NOVS) NOVARTIS AG.
 XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX Hazu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
 XX WPI: 2001-137952/14.
 DR N-PSDB: AAA89353, AAA89354.
 XX Novel gene encoding a protein that controls gene silencing, in
 PT particular silencing of plant genes
 XX Claim 1: Page 32-39: 48pp; English.
 XX The present sequence is that of Arabidopsis thaliana protein MOM,
 CC which controls gene silencing, particularly gene silencing. It
 CC is encoded by a novel silencer gene (see AAA89353) of A. thaliana
 CC ecotype Zurich. the protein exhibits partial similarity with
 CC Arpase/helicase proteins of the SWI2/SNF2 family. The sequence
 CC differs at amino acids 705 and 1219 compared with the protein
 CC (see AAB20063) encoded by the silencer gene of wild-type A.
 CC thaliana ecotype Columbia. Gene silencing is useful as a molecular
 CC tool for regulating gene expression.
 XX Sequence 2001 AA:
 SQ
 Query Match 12.8%; Score 79.5; DB 22; Length 2001;
 Best Local Similarity 21.9%; Pred. No. 30;
 Matches 25; Conservative 24; Mismatches 34; Indels 31; Gaps 4;
 QY 5 LKKEERV-----EEDNGKSDGNRGPSTVEVPTVEEVEDEFFK 43
 DB 1 MKKDEKIGLGTITVIRSLAASIPASVEQETPGLRRSSRPPSTIKVITPASATKSF--- 57
 QY 44 ILRRVHVATRTVAKVNGVAEGEJPSKKRRSQ---NLGRNSLDC-NGVRDGE 93
 DB 55 ---RLAPSPASVSKKSGGIVKNSTPSSLRSNRGKTEVSLQSSKSGSDNSIRKGD 108
 RESULT 7
 ID AAB20063 standard; Protein: 2001 AA.
 XX AAB20063:
 XX 23-APR-2001 (first entry)
 XX Arabidopsis thaliana silencing gene-encoded protein.
 XX Gene silencing; silencing gene; MOM.
 XX Arabidopsis thaliana.

[illegible]

	ID	AAG59877 standard; Protein; 252 AA.
	XX	AAG59677;
	CT	18-OCT-2000 (first entry)
	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 77458.
	KW	protein identification: signal transduction pathway; metabolic pathway;
	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
	KW	termination sequence.
	OS	Arabidopsis thaliana.
	PN	EPI0334C5-A2.
	DD	06 SEP-2000.
	FF	25-FEB-2000; 2000EP-0301439.
	PR	25-FEB-1999; 99US-0121825.
	PR	05-MAR-1999; 99US-0123180.
	PR	09-MAR-1999; 99US-0123486.
	PR	23-MAR-1999; 99US-0125788.
	PR	25-MAR-1999; 99US-0126264.
	PR	29-MAR-1999; 99US-0126785.
	PR	01-APR-1999; 99US-0127462.
	PR	06-APR-1999; 99US-0128234.
	PR	08-APR-1999; 99US-0128714.
	PR	16-APR-1999; 99US-0129845.
	PR	19-APR-1999; 99US-0130077.
	PR	21-APR-1999; 99US-0130449.
	PR	23-APR-1999; 99US-0130510.
	PR	23-APR-1999; 99US-0130891.
	PR	28-APR-1999; 99US-0131449.
	PR	30-APR-1999; 99US-0132048.
	PR	30-APR-1999; 99US-0132407.
	PR	04-MAY-1999; 99US-0132484.
	PR	05-MAY-1999; 99US-0132485.
	PR	06-MAY-1999; 99US-0132486.
	PR	06-MAY-1999; 99US-0132487.
	PR	07-MAY-1999; 99US-0132863.
	PR	11-MAY-1999; 99US-0134256.
	PR	14-MAY-1999; 99US-0134218.
	PR	14-MAY-1999; 99US-0134219.
	PR	14-MAY-1999; 99US-0134221.
	PR	14-MAY-1999; 99US-0134370.
	PR	18-MAY-1999; 99US-0134768.
	PR	19-MAY-1999; 99US-0134541.
	PR	20-MAY-1999; 99US-0135124.
	PR	21-MAY-1999; 99US-0135353.
	PR	24-MAY-1999; 99US-0135629.
	PR	25-MAY-1999; 99US-0136021.
	PR	27-MAY-1999; 99US-0136392.
	PR	28-MAY-1999; 99US-0136782.
	PR	01-JUN-1999; 99US-0137222.
	PR	03-JUN-1999; 99US-0137528.
	PR	04-JUN-1999; 99US-0137502.
	PR	07-JUN-1999; 99US-0137724.
	PR	08-JUN-1999; 99US-0138094.
	PR	10-JUN-1999; 99US-0138540.
	PR	10-JUN-1999; 99US-0138847.
	PR	14-JUN-1999; 99US-0139119.
	PR	16-JUN-1999; 99US-0139452.
	PR	16-JUN-1999; 99US-0139453.
	PR	17-JUN-1999; 99US-0139492.
	PR	18-JUN-1999; 99US-0139454.
	PR	18-JUN-1999; 99US-0139455.
	PR	18-JUN-1999; 99US-0139456.
	PR	18-JUN-1999; 99US-0139457.
	PR	18-JUN-1999; 99US-0139458.
	PR	18-JUN-1999; 99US-0139459.
	PR	18-JUN-1999; 99US-0139460.

Query Match: 12.5%; Score 76; DB 21; Length 250;

Best Local Similarity: 39.3%; Pred. No. 41;

Matches 24; Conservative 10; Mismatches 17; Indels 10; Gaps 4;

QY	28	EVKRTVTELEVDEFFKFL-----RKVVHATRIYAKYNGSVARGE.PSKKKRSQNLGLF	R:
		:	
Dd	129	EVET---KSLEDAFKLLIKQPV:ARLHVSPDLONGGVGEG-LSSIARKESRYVGIR	184
		:	
QY	82	N 82	
		:	
Dd	185	D 185	

RESULT 10 AAG59877


```
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141267.
PR 01-JUL-1999; 99US-0141362.
PR 01-JUL-1999; 99US-0142354.
PR 02-JUL-1999; 99US-0142505.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142677.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145152.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148342.
PR 13-AUG-1999; 99US-0148505.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149724.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149910.
PR 25-AUG-1999; 99US-0150556.
PR 26-AUG-1999; 99US-0150634.
PR 27-AUG-1999; 99US-0151064.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
```

```
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157751.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159364.
PR 13-OCT-1999; 99US-0159495.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 16-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162442.

Query Match 12.38; Score 76.5; DB 21; Length 230;
Best Local Similarity 29.88; Pred. No. 4;
Matches 25; Conservative 13; Mismatches 29; Indels 17; Gaps 4;

QY 2 NNSLKKEERVEEDNGKSGNRGKPESTEVYRTVTEEVTEVEFFKILRRVHVAVRTVAKVNGG 61
DB 122 NNRLLNDPDSKEE---IDNNRQKDD---RSVYEEELRE-----RLH---RKIDELKGG 154
QY 52 VAEGELPSKKRKRSONGLRNSLD 85
DB 165 RGSDDPRKSNERRKKNLNFNKKRD 188

RESULT 15
AG04594
ID AG04594 standard; Protein: 241 AA.
XX AG04594;
XX
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 819.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
```

XX OS Arab:Gopsis thaliana.
XX PN EPI033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PE 25-FEB-2000; 2C00EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123546.
PR 23-MAR-1999; 99US-0125786.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130445.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131445.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132457.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136022.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137526.
PR 04-JUN-1999; 99US-0137592.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141267.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142603.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145056.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145069.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147193.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148585.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.


```
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157885.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159437.
PR 14-OCT-1999: 99US-0159538.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 25-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 12.3%; Score 76.5; DB 21: Length 241;
Best Local Similarity 29.8%; Pred. No. 4.3;
Matches 25; Conservative 13; Mismatches 29; Indels 17; Gaps 4;
QY 2 NNSLKKEERVEEDNGKSDGNRGKSPTEVTVTTEEEVDEFFKILRVHVAVRTVAAVNGG 62
DB 133 NNRINDDDSKEE----TDNNRQKDD----RSVTYEMLRE-----RLH---RKIDELKGG 175
QY 62 VARGELFSKKRKPSONGJLKNLS:D 55
DB 176 RGSDDPRSPNSNERKKKLNLPKRRRD 199
```

Search completed: December 4, 2002, 09:22:29
Job time : 79 secs

GenCore version: 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 09:22:38 ; Search time 27 seconds
(without alignments)
132.948 Million cell updates/sec

Title: US-09-733-685-2
Perfect score: 623
Sequence: 1 MNNSLKKEERVEEDNKSUG.....QGLGLDLNCKIEPDVSLSL 122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2.6/prodata/1/aa/5A-COMB.pep:*
2: /cgn2.6/prodata/1/aa/5B-COMB.pep:*
3: /cgn2.6/prodata/1/aa/5A-COMB.pep:*
4: /cgn2.6/prodata/1/aa/5B-COMB.pep:*
5: /cgn2.6/prodata/1/aa/PC1US-COMB.pep:*
6: /cgn2.6/prodata/1/aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.5	12.9	866	US-08-651-999A-1	Sequence 1, Appl
2	80.5	12.9	866	US-09-385-752-1	Sequence 1, Appl
3	80.5	12.9	968	US-08-651-999A-7	Sequence 7, Appl
4	80.5	12.9	968	US-09-385-752-7	Sequence 7, Appl
5	71	11.4	217	US-08-971-158-1	Sequence 1, Appl
6	71	11.4	217	US-08-416-488-1	Sequence 1, Appl
7	70.5	11.3	859	US-08-053-614-2	Sequence 2, Appl
8	70.5	11.3	859	US-08-316-397B-2	Sequence 2, Appl
9	70.5	11.3	859	US-09-034-306-2	Sequence 2, Appl
10	70.5	11.3	859	US-09-259-437-2	Sequence 2, Appl
11	70.5	11.3	859	PCT-US93-09782-2	Sequence 2, Appl
12	70.5	11.3	1181	US-08-053-614-4	Sequence 4, Appl
13	70.5	11.3	1181	US-08-316-397B-4	Sequence 4, Appl
14	70.5	11.3	1181	US-09-034-306-4	Sequence 4, Appl
15	70.5	11.3	1181	US-09-259-437-4	Sequence 4, Appl
16	70.5	11.3	1181	PCT-US93-09782-4	Sequence 4, Appl
17	69	11.1	591	US-08-145-995A-21	Sequence 21, Appl
18	69	11.1	591	US-08-451-747-21	Sequence 21, Appl
19	69	11.1	591	US-09-134-852-21	Sequence 21, Appl
20	66.5	10.7	349	US-08-955-918C-2	Sequence 2, Appl
21	66.5	10.7	424	US-09-661-711A-8	Sequence 8, Appl
22	66.5	10.7	455	US-09-661-711A-4	Sequence 4, Appl
23	66.5	10.7	455	US-09-661-711A-26	Sequence 26, Appl
24	66.5	10.7	542	US-08-701-380-2	Sequence 2, Appl
25	66.5	10.7	542	US-09-032-365A-13	Sequence 13, Appl
26	66.5	10.7	551	US-09-661-711A-2	Sequence 2, Appl
27	66	10.6	251	US-08-592-214A-12	Sequence 12, Appl

28 66 10.6 251 3 US-08-659-188-12 Sequence 12, Appl
29 66 10.6 251 3 US-08-655-227-12 Sequence 12, Appl
30 66 10.6 251 3 US-08-655-241-12 Sequence 12, Appl
31 66 10.6 251 3 US-09-149-976-12 Sequence 12, Appl
32 66 10.6 251 4 US-09-398-326-12 Sequence 12, Appl
33 66 10.6 552 4 US-09-661-711A-10 Sequence 10, Appl
34 65.5 10.5 349 4 US-08-697-766A-2 Sequence 2, Appl
35 65.5 10.5 1024 4 US-09-562-737-82 Sequence 82, Appl
36 65.5 10.5 1170 4 US-09-749-588-2 Sequence 2, Appl
37 65.5 10.5 15281 2 US-08-471-119A-2 Sequence 281, Appl
38 65 10.4 180 4 US-09-615-192A-281 Sequence 4, Appl
39 65 10.4 196 3 US-07-998-289B-4 Sequence 51, Appl
40 64.5 10.4 148 1 US-08-166-195A-51 Sequence 51, Appl
41 64.5 10.4 148 2 US-08-436-883B-51 Sequence 51, Appl
42 64.5 10.4 599 1 US-08-752-238-3 Sequence 3, Appl
43 64 10.3 599 3 US-09-085-603B-3 Sequence 3, Appl
44 64 10.3 599 3 US-09-031-897-7 Sequence 7, Appl
45 64 10.3 599 3

ALIGNMENTS

RESULT 1
US-08-651-999A-1
; Sequence 1, Application US/08651999A
; Patent No. 6031086
; GENERAL INFORMATION:
; APPLICANT: Stefan Somlo and Toshio Mochizuki
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSCHILD & EHENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651.999A
FILING DATE: MAY 23, 1996
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/395
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 866 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
DESCRIPTION: NO
HYPOTHETICAL: NO
FRAGMENT TYPE: internal fragment
US-08-651-999A-1

Query Match 12.9%; Score 80.5; DB 3; Length 866;
Best Local Similarity 28.9%; Pred. No. 0.59;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
QY 8 EERVEEDNKGSDGNRGKSTEVRTVTVEEYDFKILRRV----HVATRTVAKVNGVA 63

DB 710 EDDDEDSGHSRRRGSISSGV-----SYEEFQVLRVRVDRMEHSIGSVSKIDAVI 752

QY 64 EGELPSK-KRKRSONGLRNSLDGNGV-----RDGEFDEINRVGLG-----103

DB 763 KLEIMERAKLRREVIG--RLID--GVAEDERLGRDSEIHFEQFVLVRVDRMEHSIGSVSKIDAVI 818

QY 104 -----GLG--LDLNCKPEPS 117

DB 819 ASQISHGLGTPVGLNGQPRPS 840

RESULT 2

US-09-385-752-1

; Sequence 1, Application US/09385752

; Patent No. 6228591

; GENERAL INFORMATION:

; APPLICANT: Stefan Somlo and Toshio Mochizuki

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE

; TITLE OF INVENTION: AND USES THEREOF

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

; STREET: 90 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10016

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE

; MEDIUM TYPE: DISKETTE

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09385752

; FILING DATE:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/651,993

; FILING DATE: MAY 23, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: ELIZABETH A. BOGOSIAN

; REGISTRATION NUMBER: 39,911

; REFERENCE/DOCKET NUMBER: 96700/395

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 697-5395

; TELEFAX: (212) 286-0854 or 286-0082

; TELEX: TWX 710-581-4766

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 866 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE:

; DESCRIPTION: peptide

; HYPOTHETICAL: NO

; FRAGMENT TYPE: internal fragment

US-09-385-752-1

Query Match 12.98; Score 80.5; DB 4; Length 866;

Best Local Similarity 28.98; Pred. No. 0.69;

Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;

QY 8 ERVEEDNGKSDGNGKPSSTEVTVITEEVEDEFFKILRRV-----HYAIVTVAKVNGVA 63

DB 710 EDDDEDSGHSRRRGSISSGV-----SYEEFQVLRVRVDRMEHSIGSVSKIDAVI 752

QY 64 EGELPSK-KRKRSONGLRNSLDGNGV-----RDGEFDEINRVGLG-----103

DB 763 KLEIMERAKLRREVIG--RLID--GVAEDERLGRDSEIHFEQFVLVRVDRMEHSIGSVSKIDAVI 818

QY 104 -----GLG--LDLNCKPEPS 117

DB 819 ASQISHGLGTPVGLNGQPRPS 840

RESULT 3

US-08-651-999A-7

; Sequence 7, Application US/08651999A

; Patent No. 6031089

; GENERAL INFORMATION:

; APPLICANT: Stefan Somlo and Toshio Mochizuki

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE

; TITLE OF INVENTION: AND USES THEREOF

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

; STREET: 90 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10016

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE

; MEDIUM TYPE: DISKETTE

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/651,999A

; FILING DATE: MAY 23, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: ELIZABETH A. BOGOSIAN

; REGISTRATION NUMBER: 39,911

; REFERENCE/DOCKET NUMBER: 96700/395

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 697-5395

; TELEFAX: (212) 286-0854 or 286-0082

; TELEX: TWX 710-581-4766

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 988 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE:

; DESCRIPTION: peptide

; HYPOTHETICAL: NO

US-08-651-999A-7

Query Match 12.98; Score 80.5; DB 3; Length 966;

Best Local Similarity 28.98; Pred. No. 0.69;

Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;

QY 8 ERVEEDNGKSDGNGKPSSTEVTVITEEVEDEFFKILRRV-----HYAIVTVAKVNGVA 63

DB 813 EDDDEDSGHSRRRGSISSGV-----SYEEFQVLRVRVDRMEHSIGSVSKIDAVI 865

QY 64 EGELPSK-KRKRSONGLRNSLDGNGV-----RDGEFDEINRVGLG-----103

DB 866 KLEIMERAKLRREVIG--RLID--GVAEDERLGRDSEIHFEQFVLVRVDRMEHSIGSVSKIDAVI 921

QY 104 -----GLG--LDLNCKPEPS 117

DB 922 ASQISHGLGTPVGLNGQPRPS 943

RESULT 4

US-09-385-752-7

; Sequence 7, Application US/09385752

; Patent No. 6228591

; GENERAL INFORMATION:

; APPLICANT: Stefan Somlo and Toshio Mochizuki

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE

; TITLE OF INVENTION: AND USES THEREOF

RESULT 8
US-05-316-337B-2
Sequence 2, Application US/08316397B
Patent No. 5733740
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASKO, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
APPLICANT: TOMMCROU, MURALI K.R.
TITLE OF INVENTION: THE tagA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200

```

1 ADDRESS: 127 Peachtree Street, Suite 1200
2 STREET: 127 Peachtree Street, Suite 1200
3 CITY: Atlanta
4 STATE: Georgia
5 COUNTRY: USA
6 ZIP: 30303
7
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: Floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: PatentIn Release #1.0, Version #1.25
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/06/316,357B
15 FILING DATE:
16 CLASSIFICATION: 435
17 ATTORNEY/AGENT INFORMATION:
18 NAME: SPRATT, GWENDOLYN D.
19 REGISTRATION NUMBER: 36,016
20 REFERENCE/DOCKET NUMBER: 2200, 030
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: 404/688-0770
23 TELEFAX: 404/688-9480
24 INFORMATION FOR SEQ ID NO: 2:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 859 amino acids
27 TYPE: amino acid
28 TOPOLOGY: linear
29 MOLECULE TYPE: protein
30 US-08-316-397B-2
31
32 Query Match 11.38; Score 70.5; DB 1; Length 859;
33 Best Local Similarity 24.06; Pred. No. 9;
34 Matches 31; Conservative 21; Mismatches 54; Indels 23; Gaps
35
36 ! MNKSLKKEERDEEDNGKS---DGNRCKSPTEVVRVTVEEDPEFKILRRVHVATRI-VA 56

```

Db 614 LEKSLKREHLEKDVAKNLESKGNKKNKMEAKAQAANSQKDEI--FALINKEANRARAIA 671
 QY 57 KVGNGVAEGELPSKKRKRSQNLGRN-----SLDNGVRDGEFEINRV-----GLQGL 105
 Db 672 YAQ-----NLKGIKRELSQKLENIKDKLQKSFDFGFKNGKNKQFSKAETLKALGKS 725
 QY 106 GLDLNCKPE 114
 Db 726 VKDLGINPE 734

RESULT 9

US-09-034-306-2
 ; Sequence 2, Application US/09034306
 ; Patent No. 5876943
 ; GENERAL INFORMATION:
 ; APPLICANT: COVER, TIMOTHY L.
 ; APPLICANT: BLASER, MARTIN J.
 ; APPLICANT: HARRY KLEANTHOUS
 ; APPLICANT: TUMMURU, MURALI K.R.
 ; TITLE OF INVENTION: THE tagA GENE AND METHODS FOR DETECTING
 ; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 ; STREET: 127 Peachtree Street, Suite 1200
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30303

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/034.306
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA: US/08/316.397
 APPLICATION NUMBER: US/08/316.397
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: SPRATT, GWENDOLYN D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.030
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 859 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-034-306-2
 Query Match 11.3%; Score 70.5; DB 2; Length 359;
 Best Local Similarity 24.0%; Pred. No. 9;
 Matches 31; Conservative 21; Mismatches 54; Indels 23; Gaps 5;

QY 1 MNSLKKERVEEDNGKS-----DGNRKPSDEVVTVTEEVDEFFKILRRVHVATRIVA 56
 Db 614 LEKSLKREHLEKDVAKNLESKGNKKNKMEAKAQAANSQKDEI--FALINKEANRARAIA 671
 QY 57 KVGNGVAEGELPSKKRKRSQNLGRN-----SLDNGVRDGEFEINRV-----GLQGL 105
 Db 672 YAQ-----NLKGIKRELSQKLENIKDKLQKSFDFGFKNGKNKQFSKAETLKALGKS 725
 QY 106 GLDLNCKPE 114
 Db 726 VKDLGINPE 734

RESULT 10

US-09-259-437-2
 ; Sequence 2, Application US/09259437
 ; Patent No. 6153390
 ; GENERAL INFORMATION:
 ; APPLICANT: COVER, TIMOTHY L.
 ; APPLICANT: BLASER, MARTIN J.
 ; APPLICANT: HARRY KLEANTHOUS
 ; APPLICANT: TUMMURU, MURALI K.R.
 ; TITLE OF INVENTION: THE tagA GENE AND METHODS FOR DETECTING
 ; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 ; STREET: 127 Peachtree Street, Suite 1200
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30303

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/259.437
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/316.397
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: SPRATT, GWENDOLYN D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 2200.030
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 859 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-259-437-2

Query Match 11.3%; Score 70.5; DB 4; Length 859;
 Best Local Similarity 24.0%; Pred. No. 9;
 Matches 31; Conservative 21; Mismatches 54; Indels 23; Gaps 5;

QY 1 MNSLKKERVEEDNGKS-----DGNRKPSDEVVTVTEEVDEFFKILRRVHVATRIVA 56
 Db 614 LEKSLKREHLEKDVAKNLESKGNKKNKMEAKAQAANSQKDEI--FALINKEANRARAIA 671
 QY 57 KVGNGVAEGELPSKKRKRSQNLGRN-----SLDNGVRDGEFEINRV-----GLQGL 105
 Db 672 YAQ-----NLKGIKRELSQKLENIKDKLQKSFDFGFKNGKNKQFSKAETLKALGKS 725
 QY 106 GLDLNCKPE 114
 Db 726 VKDLGINPE 734

RESULT 11

PCT-US93-09782-2
 ; Sequence 2, Application PC/TUS9309782
 ; GENERAL INFORMATION:
 ; APPLICANT: COVER, TIMOTHY L.
 ; APPLICANT: BLASER, MARTIN J.
 ; APPLICANT: TUMMURU, MURALI K.R.
 ; TITLE OF INVENTION: THE tagA GENE AND METHODS FOR DETECTING

DB 726 VKDLGINPE 734

Search completed: December 4, 2002, 05:26:48
Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen, Ltd.

QM protein - protein search, using sw model

Run on: December 4, 2002, 09:23:23 : Search time 21 seconds
(without alignments)
94,360 Million cell updates/sec

Title: US-09-733-685-2

Perfect score: 623

Sequence: 1 MNLSKXREVRVEDNKRSGD.....QGLGLDKNKPELSVLSL 122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 103943 seqs, 16242309 residues

Total number of hits satisfying chosen parameters: 103943

Minimum DB seq length: 0

Maximum DB seq length: 2500000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/us08_NEW_PUB_PEP.*
2: /cgn2_5/ptodata/2/pubpaa/PCI_NEW_PUB_PEP.*
3: /cgn2_5/ptodata/2/pubpaa/US06_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB_PEP.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB_PEP.*
7: /cgn2_6/ptodata/2/pubpaa/PCRTUS_PUBCOMB_PEP.*
8: /cgn2_5/ptodata/2/pubpaa/US06_PUBCOMB_PEP.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB_PEP.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB_PEP.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB_PEP.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB_PEP.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB_PEP.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	Description
1	80.5	12.9	866	10	US-09-753-008-1 Sequence 1, Appli
2	80.5	12.9	968	10	US-09-753-008-7 Sequence 7, Appli
3	77	12.4	966	10	US-09-628-466-6 Sequence 6, Appli
4	70	11.2	251	10	US-09-335-390A-30 Sequence 30, Appli
5	70	11.2	326	10	US-09-745-763-150 Sequence 150, Appl
6	70	11.2	323	10	US-09-325-301-969 Sequence 301, Appl
7	69.5	11.2	491	10	US-09-815-242-12023 Sequence 12023, A
8	67.5	10.8	240	10	US-09-401-369-342 Sequence 342, App
9	67.5	10.8	653	12	US-10-052-566-402 Sequence 402, App
10	66.5	10.7	159	10	US-09-815-242-5016 Sequence 5016, Ap
11	66.5	10.7	257	10	US-09-815-242-10681 Sequence 10681, A
12	66	10.6	155	10	US-09-325-301-1370 Sequence 1370, Ap
13	66	10.6	390	10	US-09-815-242-10683 Sequence 10683, A
14	65.5	10.5	1170	12	US-10-135-687-2 Sequence 2, Appli
15	65.5	10.5	1210	10	US-09-860-352A-2 Sequence 2, Appli
16	65	10.4	89	10	US-09-864-761-36753 Sequence 36753, A
17	64.5	10.4	162	10	US-09-864-761-33534 Sequence 33534, A
18	64.5	10.4	712	10	US-09-815-242-10930 Sequence 10930, A
19	64.5	10.4	2568	10	US-09-866-108-3 Sequence 3, Appli

Sequence 18, Appli
Sequence 5, Appli
Sequence 4, Appli
Sequence 312, App
Sequence 188, App
Sequence 3, Appli
Sequence 3, Appli
Sequence 161, App
Sequence 2, Appli
Sequence 1294, Ap
Sequence 10, Appl
Sequence 4, Appli
Sequence 2041, A
Sequence 7, Appli
Sequence 14, Appl
Sequence 24, Appl
Sequence 6, Appli
Sequence 13261, A
Sequence 4, Appli
Sequence 13765, A
Sequence 234, App
Sequence 34, Appl
Sequence 82, Appl
Sequence 70, Appl
Sequence 4, Appli
Sequence 12, Appli

ALIGNMENTS

RESULT 1
US-09-753-008-1
Sequence 1, Application US/09753008
Patent No. US20020051520A1
GENERAL INFORMATION:
APPLICANT: Stefan Som-O and Toshio Mochizuki
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
AND USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER. ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/753,008
FILING DATE: 02-Jan-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/385,752
FILING DATE: 1999-08-30

APPLICATION NUMBER: 08/651,999
FILING DATE: MAY 23, 1996

ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/395

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 866 amino acids

TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: <Unknown>
DESCRIPTION: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal fragment
SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-753-C08-1

Query Match 12.9%; Score 80.5; DB 10; Length 866;
Best Local Similarity 28.9%; Pred. No. 1;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
QY 8 EERVEEDNGKSDGNRGKPTSEVYRVITVEEVEDEFFKILRRV----HVATRTVAKYNGVA 63
DB 710 EEDDEDGSHSSRRGSISSGV-----SYEEFQVLVRRVDRMEHSGISVSKIDAVIV 762
QY 64 EGELPSK-KKKRSQNLGLRSLDCNGV----RDGEFDEINRVGLQ-----103
DB 763 KLEIMRAKLRREVJG--RLLD--GVAFDERLGRDSEIHREQMERLVREELERWESUDA 818
QY 104 -----GLG--LDLCKPEPDS 117
DB 819 ASOISHGGLGTPVGLNQPRRS 840

RESULT 2

US-09-753-C08-7
; Sequence 7, Application US/09753068
; Patent No. US20020061520A1
; GENERAL INFORMATION:
APPLICANT: Stefan Somlo and Toshio Mochizuki
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
AND USES THEREOF

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
DISKETTE

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/753,008

FILING DATE: 02-Jan-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/385,752

FILING DATE: 1999-08-30

APPLICATION NUMBER: 08/651,995

FILING DATE: MAY 23, 1996

ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. ROGOSIAN

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/395

TELEPHONE: (212) 697-5995

TELEFAX: (212) 296-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 7

SEQUENCE CHARACTERISTICS:

LENGTH: 968 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: <Unknown>

DESCRIPTION: peptide

HYPOTHETICAL: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 7
US-09-753-C08-7

Query Match 12.9%; Score 80.5; DB 10; Length 968;
Best Local Similarity 28.9%; Pred. No. 1.2;
Matches 41; Conservative 19; Mismatches 39; Indels 43; Gaps 8;
QY 8 EERVEEDNGKSDGNRGKPTSEVYRVITVEEVEDEFFKILRRV----HVATRTVAKYNGVA 63
DB 813 EEDDEDGSHSSRRGSISSGV-----SYEEFQVLVRRVDRMEHSGISVSKIDAVIV 865
QY 64 EGELPSK-KKKRSQNLGLRSLDCNGV----RDGEFDEINRVGLQ-----103
DB 866 KLEIMRAKLRREVJG--RLLD--GVAFDERLGRDSEIHREQMERLVREELERWESUDA 921
QY 104 -----GLG--LDLCKPEPDS 117
DB 922 ASOISHGGLGTPVGLNQPRRS 943

RESULT 3

US-09-828-466-6
; Sequence 6, Application US/09828466
; Patent No. US20020035056A1
; GENERAL INFORMATION:
APPLICANT: Curtis, Rory A.J.
TITLE OF INVENTION: 54420, A NOVEL HUMAN CALCIUM CHANNEL
FILE REFERENCE: MNI-125CP
CURRENT APPLICATION NUMBER: US/09/828,466
CURRENT FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: US 09/544,797
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 6
LENGTH: 966
TYPE: PRT
ORGANISM: Homo sapiens
US-09-828-466-6

Query Match 12.4%; Score 77; DB 10; Length 966;
Best Local Similarity 29.1%; Pred. No. 2.8;
Matches 30; Conservative 17; Mismatches 44; Indels 12; Gaps 3;

QY 8 EERVEEDNGKSDGNRGKPTSEVYRVITVEEVEDEFFKILRRV----HVATRTVAKYNGVA 53
DB 811 EEDDEDGSHSSRRGSISSGV-----SYEEFQVLVRRVDRMEHSGISVSKIDAVIV 863
QY 64 EGELPSK-KKKRSQNLGLRSLDCNGVVRDGEFDEINRVGLQGL 105
DB 864 KLEIMRAKLRREVJGRLDGVDAEDARLGRDSEIHREQMERL 906

RESULT 4

US-09-635-390A-30
; Sequence 30, Application US/09035390A
; Patent No. US20020076761A1
; GENERAL INFORMATION:
APPLICANT: Escobedo, Jaime
Oulanjin, Hu
Garcia, Pablo
Williams, Lewis T.
Kothakota, Srinivas
TITLE OF INVENTION: Secreted Human Proteins
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916

```

1  FILING DATE: 18-JUL-2000
2  CLASSIFICATION: <Unknown>
3  ATTORNEY/AGENT INFORMATION:
4  NAME: Springer, Suzanne A.
5  REGISTRATION NUMBERS: 41,323
6  TELECOMMUNICATION INFORMATION:
7  TELEPHONE: (617) 498-8284
8  TELEFAX: (617) 876-5851
9  INFORMATION FOR SEQ ID NO: 150:
10  SEQUENCE CHARACTERISTICS:
11  LENGTH: 320 amino acids
12  TYPE: amino acid
13  STRANDEDNESS: <Unknown>
14  TOPOLOGY: linear
15  MOLECULE TYPE: protein
16  SEQUENCE DESCRIPTION: SEQ ID NO: 150:
17  US-09-745-763-150
18
19  Query Match 11.2%; Score 70; DB 10; Length 320;
20  Best Local Similarity 26.1%; Pred.No. 3.6;
21  Matches 30; Conservative 17; Mismatches 26; Indels 42; Gaps
22
23  QY 25 PSTEV-----VSTVTEEVDE---FFKILRVHVHVAITVAKYNGVAGVSGELPSKKRKRSQN 77
24  Db 55 PPKEVINGNIKTVTEYKIDGKKFKIVRTFRIETRKASK--AVA-----RRKNWKK 104
25
26  QY 78 LGLRSLDCNGVRGGEF-----DINRVGLGLOGLDLNCCKPDPDSVS 119
27  Db 105 FG-----NSEFDPGPVWATTVSDVSMTFITTSKE-DLNCQEEEDPMN 147
28
29  RESULT 6
30  US-09-925-301-969
31  Sequence 969, Application US/09925301
32  Patent No. US20020052308A1
33  GENERAL INFORMATION:
34  APPLICANT: Rosen et al.
35  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
36  FILE REFERENCE: PA106
37  CURRENT APPLICATION NUMBER: US/09/925,301
38  CURRENT FILING DATE: 2001-08-10
39  PRIOR APPLICATION NUMBER: PCT/US00/05882
40  PRIOR FILING DATE: 2000-03-09
41  PRIOR APPLICATION NUMBER: 50/124,270
42  PRIOR FILING DATE: 1999-03-12
43  NUMBER OF SEQ ID NOS: 1694
44  SOFTWARE: PatentIn Ver. 2.0
45  SEQ ID NO 969
46  LENGTH: 325
47  TYPE: PRT
48  ORGANISM: Homo sapiens
49  US-09-925-301-969
50
51  Query Match 11.2%; Score 70; DB 10; Length 325;
52  Best Local Similarity 26.1%; Pred.No. 3.7;
53  Matches 30; Conservative 17; Mismatches 26; Indels 42; Gaps
54
55  QY 25 PSTEV-----VSTVTEEVDE---FFKILRVHVHVAITVAKYNGVAGVSGELPSKKRKRSQN 77
56  Db 60 PPKEVINGNIKTVTEYKIDGKKFKIVRTFRIETRKASK--AVA-----RRKNWKK 109
57
58  QY 78 LGLRSLDCNGVRGGEF-----DINRVGLGLOGLDLNCCKPDPDSVS 119
59  Db 110 FG-----NSEFDPGPVWATTVSDVSMTFITTSKE-DLNCQEEEDPMN 152
60
61  RESULT 7
62  US-09-815-242-12023
63  Sequence 12023, Application US/09815242
64  Patent No. US20020061569A1
65  GENERAL INFORMATION:
66  APPLICANT: Haselbeck, Robert
67  APPLICANT: Ohlsen, Kari L.

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

1 PRIOR APPLICATION NUMBER: 60/066772
2 PRIOR FILING DATE: 1997-11-24
3 PRIOR APPLICATION NUMBER: 60/069335
4 PRIOR FILING DATE: 1997-12-11
5 PRIOR APPLICATION NUMBER: 60/069425
6 PRIOR FILING DATE: 1997-12-12
7 PRIOR APPLICATION NUMBER: 60/069670
8 PRIOR FILING DATE: 1997-12-17
9 PRIOR APPLICATION NUMBER: 60/068317
10 PRIOR FILING DATE: 1997-12-18
11 PRIOR APPLICATION NUMBER: 60/077450
12 PRIOR FILING DATE: 1998-03-10
13 PRIOR APPLICATION NUMBER: 60/077632
14 PRIOR FILING DATE: 1998-03-11
15 PRIOR APPLICATION NUMBER: 60/077649
16 PRIOR FILING DATE: 1998-03-11
17 PRIOR APPLICATION NUMBER: 60/078886
18 PRIOR FILING DATE: 1998-04-20
19 PRIOR APPLICATION NUMBER: 60/078939
20 PRIOR FILING DATE: 1998-03-20
21 PRIOR APPLICATION NUMBER: 60/079664
22 PRIOR FILING DATE: 1998-03-27
23 PRIOR APPLICATION NUMBER: 60/079786
24 PRIOR FILING DATE: 1998-03-27
25 PRIOR APPLICATION NUMBER: 60/080107
26 PRIOR FILING DATE: 1998-03-31
27 PRIOR APPLICATION NUMBER: 60/080194
28 PRIOR FILING DATE: 1998-03-31
29 PRIOR APPLICATION NUMBER: 60/080327
30 PRIOR FILING DATE: 1998-04-01
31 PRIOR APPLICATION NUMBER: 60/080333
32 PRIOR FILING DATE: 1998-04-01
33 PRIOR APPLICATION NUMBER: 60/081049
34 PRIOR FILING DATE: 1998-04-08
35 PRIOR APPLICATION NUMBER: 60/081070
36 PRIOR FILING DATE: 1998-04-08
37 PRIOR APPLICATION NUMBER: 60/081195
38 PRIOR FILING DATE: 1998-04-09
39 PRIOR APPLICATION NUMBER: 60/081838
40 PRIOR FILING DATE: 1998-04-15
41 PRIOR APPLICATION NUMBER: 60/082568
42 PRIOR FILING DATE: 1998-04-21
43 PRIOR APPLICATION NUMBER: 60/082568
44 PRIOR FILING DATE: 1998-04-21
45 PRIOR APPLICATION NUMBER: 60/082704
46 PRIOR FILING DATE: 1998-04-22
47 PRIOR APPLICATION NUMBER: 60/082797
48 PRIOR FILING DATE: 1998-04-22
49 PRIOR APPLICATION NUMBER: 60/083322
50 PRIOR FILING DATE: 1998-04-28
51 PRIOR APPLICATION NUMBER: 60/083495
52 PRIOR FILING DATE: 1998-04-29
53 PRIOR APPLICATION NUMBER: 60/083496
54 PRIOR FILING DATE: 1998-04-29
55 PRIOR APPLICATION NUMBER: 60/083499
56 PRIOR FILING DATE: 1998-04-29
57 PRIOR APPLICATION NUMBER: 60/083559
58 PRIOR FILING DATE: 1998-04-29
59 PRIOR APPLICATION NUMBER: 60/084366
60 PRIOR FILING DATE: 1998-05-05
61 PRIOR APPLICATION NUMBER: 60/084414
62 PRIOR FILING DATE: 1998-05-06
63 PRIOR APPLICATION NUMBER: 60/084635
64 PRIOR FILING DATE: 1998-05-07
65 PRIOR APPLICATION NUMBER: 60/084640
66 PRIOR FILING DATE: 1998-05-07
67 PRIOR APPLICATION NUMBER: 60/084643
68 PRIOR FILING DATE: 1998-05-07
69 PRIOR APPLICATION NUMBER: 60/085573
70 PRIOR FILING DATE: 1998-05-15
71 PRIOR APPLICATION NUMBER: 60/085579
72 PRIOR FILING DATE: 1998-05-15
73 PRIOR APPLICATION NUMBER: 60/085580

74 PRIOR FILING DATE: 1998-05-15
75 PRIOR APPLICATION NUMBER: 60/085582
76 PRIOR FILING DATE: 1998-05-15
77 PRIOR APPLICATION NUMBER: 60/085700
78 PRIOR FILING DATE: 1998-05-15
79 PRIOR APPLICATION NUMBER: 60/086023
80 PRIOR FILING DATE: 1998-05-18
81 PRIOR APPLICATION NUMBER: 60/086392
82 PRIOR FILING DATE: 1998-05-22
83 PRIOR APPLICATION NUMBER: 60/086486
84 PRIOR FILING DATE: 1998-05-22
85 PRIOR APPLICATION NUMBER: 60/087098
86 PRIOR FILING DATE: 1998-05-28
87 PRIOR APPLICATION NUMBER: 60/087208
88 PRIOR FILING DATE: 1998-05-28
89 PRIOR APPLICATION NUMBER: 60/087609
90 PRIOR FILING DATE: 1998-06-02
91 PRIOR APPLICATION NUMBER: 60/087759
92 PRIOR FILING DATE: 1998-06-02
93 PRIOR APPLICATION NUMBER: 60/087827
94 PRIOR FILING DATE: 1998-06-03
95 PRIOR APPLICATION NUMBER: 60/088025
96 PRIOR FILING DATE: 1998-06-04
97 PRIOR APPLICATION NUMBER: 60/088028
98 PRIOR FILING DATE: 1998-06-04
99 PRIOR APPLICATION NUMBER: 60/088029
100 PRIOR FILING DATE: 1998-06-04
101 PRIOR APPLICATION NUMBER: 60/088033
102 PRIOR FILING DATE: 1998-06-04
103 PRIOR APPLICATION NUMBER: 60/088167
104 PRIOR FILING DATE: 1998-06-05
105 PRIOR APPLICATION NUMBER: 60/088202
106 PRIOR FILING DATE: 1998-06-05
107 PRIOR APPLICATION NUMBER: 60/088212
108 PRIOR FILING DATE: 1998-06-05
109 PRIOR APPLICATION NUMBER: 60/088217
110 PRIOR FILING DATE: 1998-06-05
111 PRIOR APPLICATION NUMBER: 60/088326
112 PRIOR FILING DATE: 1998-06-04
113 PRIOR APPLICATION NUMBER: 60/088655
114 PRIOR FILING DATE: 1998-06-09
115 PRIOR APPLICATION NUMBER: 60/088722
116 PRIOR FILING DATE: 1998-06-10
117 PRIOR APPLICATION NUMBER: 60/088736
118 PRIOR FILING DATE: 1998-06-10
119 PRIOR APPLICATION NUMBER: 60/088740
120 PRIOR FILING DATE: 1998-06-10
121 PRIOR APPLICATION NUMBER: 60/088811
122 PRIOR FILING DATE: 1998-06-10
123 PRIOR APPLICATION NUMBER: 60/088824
124 PRIOR FILING DATE: 1998-06-10
125 PRIOR APPLICATION NUMBER: 60/088825
126 PRIOR FILING DATE: 1998-06-10
127 PRIOR APPLICATION NUMBER: 60/088826
128 PRIOR FILING DATE: 1998-06-10
129 PRIOR APPLICATION NUMBER: 60/088861
130 PRIOR FILING DATE: 1998-06-11
131 PRIOR APPLICATION NUMBER: 60/088863
132 PRIOR FILING DATE: 1998-06-11
133 PRIOR APPLICATION NUMBER: 60/088876
134 PRIOR FILING DATE: 1998-06-11
135 PRIOR APPLICATION NUMBER: 60/089090
136 PRIOR FILING DATE: 1998-06-12
137 PRIOR APPLICATION NUMBER: 60/089105
138 PRIOR FILING DATE: 1998-06-12
139 PRIOR APPLICATION NUMBER: 60/089512
140 PRIOR FILING DATE: 1998-06-16
141 PRIOR APPLICATION NUMBER: 60/089514
142 PRIOR FILING DATE: 1998-06-16
143 PRIOR APPLICATION NUMBER: 60/089538
144 PRIOR FILING DATE: 1998-06-17
145 PRIOR APPLICATION NUMBER: 60/089598
146 PRIOR FILING DATE: 1998-06-17

;; PRIOR APPLICATION NUMBER: 60/089053
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089058

Query Match 10.8%; Score 67.5; DB 12; Length 653;

Best Local Similarity 23.9%; Pred. No. 17;
Matches 32; Conservative 18; Mismatches 55; Indels 29; Gaps 4;

Qy 1 MNLSKKEEVEEDKSGDNKPKSTEVVTVTEVEDEFFKILRR-----VHVATR 53

Db 429 MOKSLERLREKAEKEEKQTKK--ETSAELRLKLKEE--KRLKKKRRKSISSSSVSSAUE 486

Qy 54 TVAKVNGVAGELPSKKRKRKSNGLRNS-----LDNGVGDGDFEINRVGLQGL 107

Db 487 SVSSSSSSSGHKRHKKRRKRSSESSRRHSSRRSSNQIDQNRKDE----- 535

Qy 108 DLNCKPPDSVSL 121

Db 536 ---CYPVPANTSAS 546

RESULT 10

US-09-615-242-5016

Sequence 5016, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert I.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,846

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 1411C

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5016

LENGTH: 159

TYPE: PRT

ORGANISM: Enterococcus faecalis

US-09-815-242-5015

Query Match 10.7%; Score 66.5; DB 10; Length 159;

Best Local Similarity 28.9%; Pred. No. 3.4;
Matches 26; Conservative 11; Mismatches 40; Indels 13; Gaps 3;

Qy 19 DGNRCKPSTEVVTVTEVEDEFFKIL--RRVHVATRVAKYNGVAGELPSKKRKS 75

Db 75 DGNMVKPRLVETIETINEEGTSVVRVLIQEGKFGVKNRLATVGKPV-----YLRKLRM 129

Qy 76 QNLGLRNSLDGNGVRDGEFDEINRVGLQL 105

Db 130 GELNLDLTEL-----GEYRPLTEARLQQL 154

RESULT 11

US-09-815-242-10681

Sequence 10681, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert I.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,846

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 1411C

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10681

LENGTH: 257

TYPE: PRT

ORGANISM: Enterococcus faecalis

US-09-815-242-10681

Query Match 10.7%; Score 66.5; DB 10; Length 257;

Best Local Similarity 28.9%; Pred. No. 6.4;

Matches 26; Conservative 11; Mismatches 40; Indels 13; Gaps 3;

Qy 19 DGNRCKPSTEVVTVTEVEDEFFKIL--RRVHVATRVAKYNGVAGELPSKKRKS 75

Db 130 DGNMVKPRLVETIETINEEGTSVVRVLIQEGKFGVKNRLATVGKPV-----YLRKLRM 224

Qy 76 QNLGLRNSLDGNGVRDGEFDEINRVGLQL 105

Db 225 GELNLDLTEL-----GEYRPLTEARLQQL 249

RESULT 12

US-09-925-301-1370

Sequence 1370, Application US/09925301

Patent No. US20020052308A1

GENERAL INFORMATION:

APPLICANT: Rosen, et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA106

CURRENT APPLICATION NUMBER: US/09/925,301

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05882

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1694

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1370

LENGTH: 155

TYPE: PRT

```

1  ORGANISM: Homo sapiens
US-09-925-301-1370

Query Match      10.6%; Score 66; DB 10; Length 155;
Best Local Similarity 32.3%; Pred. No. 37;
Matches 20; Conservative 14; Mismatches 24; Indels 4; Gaps 0;

CY 42 FKLRRVHVATVAVKVG-GVAEGELPSKKRKRKSNGLRNSLDCNGVKGGEDEHNR 89
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11 FPIARQEE..SVTTIAVMSGRGQGGKAKAKAKRSRSBAGLQFFVGVGIALKQKNAE-R 68

CY 100 VG 101
      ||
DB 69 VG 70

RESULT 13
US-09-815-242-10683
Sequence 10683, Application US/098/5242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Walli, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: FLITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/151,678
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,579
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,304
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10683
LENGTH: 350
TYPE: PRT
ORGANISM: Enterococcus faecalis

JS-09-815-242-10683

Query Match      10.6%; Score 66; DB 10; Length 350;
Best Local Similarity 28.1%; Pred. No. 13;
Matches 34; Conservative 18; Mismatches 49; Indels 23; Gaps 0;

JY 3 NSLKKERVEEDNCKSDGN----NGKPSITVWTVTVEEVDEFFK --FLRRVHVATRV 55
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
JB 56 DSRRSRRPHQTDGSGNGSDQPPRGKKDKRKKRKKSKTKRFRKQV..TILILFAVSTV 115

JY 56 AKVNGGVA---EGELPSKKRK-----RSCNGLRNSL----DCNGVPGGEFTGLNAVGLQ 103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
JB 116 MFLKGSAAEHCDSPGEKVETTFNGVKSSN-GAKNILLIGSDTGTGDFGADRAFTIMVQLEN 174

JY 104 G 104
      |
JB 175 G 175

RESULT 14
US-10-135-687-2
Sequence 2, Application US/10135687
Patent No. US20020123120A1
GENERAL INFORMATION:
APPLICANT: CHANDRAMOULI SWARAN, Ishwar et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL001068DV
CURRENT APPLICATION NUMBER: US/10/135,687
CURRENT FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: 09/749,588
PRIOR FILING DATE: 2000-12-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1170
TYPE: PRT
ORGANISM: Homo sapiens
US-10-135-687-2

Query Match      10.5%; Score 65.5; DB 12; Length 1170;
Best Local Similarity 38.6%; Pred. No. 62;
Matches 17; Conservative 5; Mismatches 15; Indels 7; Gaps 1;

CY 49 HVATRVAVKVGVA-----EGELPSKKRKRKSNGLRNSLD 85
      -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -|
DB 776 HVTLATAPLVNGVAHVYVROOQSSSLPSKKKKNKQSAFVSSKSSLD 819

RESULT 15
US-09-860-352A-2
Sequence 2, Application US/09860352A
Patent No. US20020132789A1
GENERAL INFORMATION:
APPLICANT: Millenium Pharmaceuticals, Inc.
APPLICANT: Curtis, Rory
APPLICANT: Weich, Nading
TITLE OF INVENTION: 13305 NOVEL PROTEIN KINASE MOLECULES AND
TITLE OF INVENTION: USES THEREFOR
FILE REFERENCE: 38155-20016.00
CURRENT APPLICATION NUMBER: US/09/860,352A
CURRENT FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: US 60/205,301
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1210
TYPE: PRT
ORGANISM: Homo sapiens
US-09-860-352A-2

Query Match      10.5%; Score 65.5; DB 10; Length 1210;
Best Local Similarity 38.6%; Pred. No. 65;
Matches 17; Conservative 5; Mismatches 15; Indels 7; Gaps 1;

CY 49 HVATRVAVKVGVA-----EGELPSKKRKRKSNGLRNSLD 85
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 816 HVTLATAPLVNGVAHVYVROOQSSSLPSKKKKNKQSAFVSSKSSLD 859

Search completed: December 4, 2002, 09:27:28
Job time : 22 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compaq Inc.

OM protein - protein search, using sw model

Run on: December 4, 2002, 09:21:19 : Search time 44 seconds
(without alignments)
266,555 Million cells updates/sec

Title: US-09-733-685-2

Perfect score: 623

Sequence: 1 MNNSLKKEERVEDNGKSDG.....GSGELGLDLMKKKRDGVSNSL 122

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 6.5

Searched: 283224 seqs, 9633422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY-ES

Result No.	Score	Query Match	Length	DB ID	Description
1	80.5	12.9	265	2 A86548	hypothetical prote
2	80.5	12.9	283	2 H72075	hypothetical prote
3	80.5	12.9	604	2 G02640	polycystic kidney
4	80.5	12.9	2254	2 D86215	protein T5022.14
5	78.5	12.6	245	2 C81592	hypothetical prote
6	76	12.2	1007	2 H81670	conserved hypote
7	74	11.9	141	2 T00928	hypothetical prote
8	74	11.9	307	2 G69211	conserved hypote
9	74	11.9	891	2 G96436	hypothetical prote
10	73.5	11.8	389	2 C86450	P14M2.18 protein
11	73	11.7	107	2 C81232	50S ribosomal prot
12	73	11.7	350	2 T48522	hypothetical prote
13	73	11.7	2149	2 T47555	hypothetical prote
14	72.5	11.6	448	2 S63961	hypothetical prote
15	72.5	11.6	484	2 H50140	transcription fact
16	72.5	11.6	804	2 T32864	hypothetical prote
17	72	11.6	132	2 T50108	yeast ARG2 protei
18	71.5	11.5	335	2 D83142	hypothetical prote
19	71.5	11.5	543	2 S53817	threosome beta sh
20	71.5	11.5	2275	2 T33123	hypothetical prote
21	71	11.4	467	2 D66485	protein F28J9.13
22	71	11.4	862	2 T01141	hypothetical prote
23	70.5	11.3	302	1 TPCHRC	treponem T. cardia
24	70.5	11.3	314	2 T29363	hypothetical prote
25	70.5	11.3	775	1 WMBE19	ribonucleoside-dip
26	70.5	11.3	2297	2 AB2494	hypothetical prote
27	70	11.2	227	2 S57480	mclbdopterin-guan
28	70	11.2	853	2 T23697	hypothetical prote
29	69.5	11.2	255	2 D71404	hypothetical prote

ALIGNMENTS

RESULT 1

A86548

hypothetical protein CPJ0461 [imported] - Chlamydia pneumoniae (strain J138)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001

C:Accession: A86548

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: A86548

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <STO>

A:Cross-references: GB:BA000008; NID:g8978831; PID:BAA38667.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0461

Query Match 12.9% Score 80.5; DB 2; Length 265;

Best Local Similarity 27.2%; Pred. NC. 3.9;

Matches 40; Conservative 19; Mismatches 37; Indels 51; Gaps 7;

QY 12 EFQNGKSDGNHGK-----PSIEVVRVTVEE---VDEFFKILR-RVYVATRVAKVNS 62

DR 104 QEDGKVGGRTPEDPVDITPTIEVQLLPDEELSTVDEALQGVRSNLTITAYRSVEK---- 159

QY 62 VAEGELPSKKRKRSONGLRNSLD-----CNGVR-----DGER 94

DR 160 -----PMIIDLAVGEGLRDSABLNEFRLANGVQNHYPHTVKLYLAKNLADVWDECR 213

QY 95 DFNRVGTGGLGLO:NCKPPEDSVLS 121

DR 214 SEERKGLRALGLD-----PRTESISLT 236

RESULT 2

H72075

hypothetical protein - Chlamydia pneumoniae (strain CWL029)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C:Accession: H72075

R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, J.; Grimwood, Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: H72075

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <ARN>

A:Cross-references: GB:AE001630; GB:AE001363; NID:g4376740; PID:RAD18603.1; PID:g437

A:Experimental source: strain CWL029

```
C;Genetics:
A:Gene: CPh0461

Query Match      12.9%; Score 80.5; DB 2; Length 265;
Best Local Similarity 27.2%; Pred. No. 3.9;
Matches 40; Conservative 19; Mismatches 37; Indels 51; Gaps 7;

QY 12 EDNGKSDGNRGK-----PSTEVVTVTEEE---VDEFFKILR-RVHVATRTVAKVNGG 61
   ||| ||| : ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
Db 104 QEDLGRVHGRTPEDPVDITPTEIVOLLPEDEISTVDEALQGVRSRLTYVRSVEK---- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 52 VAGSELPKSKRKRSQNLGLRNSLD-----CNGVR-----DGEFF 94
   : : : : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 160 -----PMIQDLALVGFGLRSDADLINFVRLANGVQNHYPHTKVKLYLAKN-ADVWDCEI 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 95 DEINRVGJQGLGLDNCPEPDSVLS 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 SEEEKQLRALGLD----PKIESISLT 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
G02640
polycystic kidney disease protein 2 - human (fragment)
N:Alternate names: polycystin
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C:Accession: G02640
R:Schneider, K.C.
submitted to the EMBL Data Library, April 1996
A:Reference number: H01518
A:Accession: G02640
A:Status: translated from GB/EMS1/CBBJ
A:Molecule type: mRNA
A:Residues: 1-608 <SCH>
A:Cross-references: EMBL:U56813; NID:g147759; PIDN:AA50933.1; PID:g14776-2
C:Genetics:
A:Gene: GDB:PKD2; PKD4
A:Cross-references: GDB:118651; OMIM:173510
A:Map position: 4q21-4q23

Query Match      12.9%; Score 80.5; DB 2; Length 608;
Best Local Similarity 28.9%; Pred. No. 9.8;
Matches 41; Conservative 19; Mismatches 19; Indels 43; Gaps 6;

QY 8 EERVEENKSGDGNRKPSTEVVTVTEEEVDEFFKILRV-----HVATRTVAKVNGGVA 63
   ||| ||| : ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
Db 453 SEEDDEDSGHSRRKSGSISSGV-----SYEEFQVIVRVDSMHSISGYSKILGAVTV 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 64 EGELPSK-KRKRSQNLGRNSLDNGV-----DGEFFDSEINRVSLG----- 233
   : : : : : ||| ||| : ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
Db 506 KLEINERAKLKRREVLG--RLLD--GVAEUER-GRJSEIHKEGNERLVRFELERWESDVA 561
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 104 -----GIG--LDLNCPEPDS 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 562 ASQISHGLGTFVGLNCPRRPS 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
G06215
protein T6D22.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D86215
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, G.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Kraykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maizels, R.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
```

```
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86215
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2254 <STO>
A:Cross-references: GB:AE005172; NID:g8778840; PIDN:AAF79839.1; GSPDB:GX00141
C:Genetics:
A:Gene: T6D22.14
A:Map position: 1

Query Match      12.9%; Score 80.5; DB 2; Length 2254;
Best Local Similarity 21.7%; Pred. No. 42;
Matches 25; Conservative 25; Mismatches 34; Indels 31; Gaps 4;

QY 4 SLKKEERV-----EEDNGKSDGNRKPSTEVVTVTEEEVDEFF 42
   : ||| ||| : ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
Db 232 NMKDEKIGLTGRTIYTRSLAASIPASVEQETPCGLRRSRGTPSTKVTIPASATRKSE-- 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 43 KILRRVHVATRTVAKVNGVAGSELPKSKRKRSQ---NLGRNSLDC-NGVRDGE 93
   : : : : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 290 -----RLAPSPASVSKSGGVKNSVTFSSLRSRNRGKTEVSLQSSKSGSDNSIRKGD 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
C81592
hypothetical protein CP0291 [imported] - Chlamydia pneumoniae (strain AR39)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: C81592
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
rger, S.; Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: C81592
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <REA>
A:Cross-references: GB:AE002191; GB:AE002161; NID:g7189216; PIDN:AAF38144.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0291

Query Match      12.6%; Score 78.5; DB 2; Length 265;
Best Local Similarity 27.4%; Pred. No. 6;
Matches 40; Conservative 18; Mismatches 37; Indels 51; Gaps 7;

QY 13 EDNGKSDGNRGK-----PSTEVVTVTEEE---VDEFFKILR-RVHVATRTVAKVNGGV 62
   : ||| : : ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
Db 105 EDCOKVHGRTPEDPVDITPTEIVOLLPEDEISTVDEALQGVRSRLTYVRSVEK----- 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 63 AFGEPLSKKKRKRSQNLGLRNSLD-----CNGVR-----DGEFF 95
   : : : : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 160 -----PMIQDLALVGFGLRSDADLINFVRLANGVQNHYPHTKVKLYLAKNADVWDCEIS 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 96 EINRVGJQGLGLDNCPEPDSVLS 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 BEEKQLRALGLD----PKIESISLT 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
H81670
conserved hypothetical protein TC0741 [imported] - Chlamydia muridarum (strain Nig)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: H81670
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
rger, S.; Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: H81670
A:Status: preliminary
```

```

C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: G59211
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, R.; Dubois, C.; Aldredge, T.
; Liu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani,
; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
; Bartolot, I. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A59000; MUID:98037514; PMID:9371463
A:Accession: G59211
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 5397 <MTE>
A:Cross-references: GB:AE000660; GB:AE000666; NID:g2621015; PIDN:AAB85336.1; P.D:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH838

Query Match 11.9% Score 74: DB 2: Length 307:
Best Local Similarity 26.9% Pred. No. 18:
Matches 32: Conservative 13: Mismatches 38: Indels 36: Gaps 4:

QY 5 LKKEERVEEDNGKSDGN-----RGKDPSTEVVTVTVEEYDEFFKI----- 44
DB 3 LKQ:ERKTSQSPSDT.LRVKGLVINGKPIAEAIRSVKGESEETWEISEKESHGDRYD 52
QY 45 ---LRRVIVATRVAKVGVGVGELPSKKRKRKSNLGRNSLDCNGVRDGEFEINRV 100
DB 63 VEKLRIIRAOGRSLRNQGSTIE----RLKRERKSLLGKIRELE-----DEXSL 103

RESULT 9
C96636
Hypothetical protein Flp17.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96636
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Aton
; Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.A.; Devar,
; Nuss, D.; Kierman, W.C.; White, G.; Eisen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, J.
; Nature 406, 8-6-820, 2000
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis.
A:Reference number: A86441; MUID:21016719; PMID:11130712
A:Accession: G96636
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-691 <STO>
A:Cross-references: GB:AE005173; NID:g2443891; PIDN:AAB71484.1; GSPDB:GN06141
C:Genetics:
A:Gene: Flp17.17
A:Map position: 1

Query Match 11.9% Score 74: DB 2: Length 891:
Best Local Similarity 28.6% Pred. No. 60:
Matches 22: Conservative 16: Mismatches 33: Indels 6: Gaps 2:

QY 2 NNSLKKSERVEEDNGKSDGNRGKDPSTEVVTVTVEEYDEFFKILRRVHVATRVAKVNGG 61
DB 784 SNGTAQEK--EDKSGEKKNNVKNPETHKSKTSPTRATETMSKTAKKPTVASRMAQKNKF 841
QY 62 VAEGELPSKKRKRKSNL 78
DB 842 EKEEEM----KKRIENL 854

RESULT 10
C86460
Fl4M2.1a protein - Arabidopsis thaliana

```

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov 2001
C:Accession: C86460
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, C.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.R.; Dewar, K.;
Nansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; M01D:21016719; PMID:11130712
A:Accession: C86460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STO>
A:Cross-references: GB:AE005172; NID:g9665104; P1CN:AAF57295.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
Query Match 11.8%; Score 73.5; DB 2; Length 189;
Best Local Similarity 28.3%; Pred. No. 27;
Matches 30; Conservative 17; Mismatches 38; Indels 21; Gaps 6;
QY 7 KEERVELDNCKSGNR-GKPSSTFVR-TVTEFEVDFKILRRVHVVAIRTVAKVNGGV 62
DB 82 KEETKPTAQSDENQGNPLIEKQVSTIDAESD-----KKMEDNESKEVNISSQNE 135
QY 65 GELPSKK-----KKRSNGLNLSLDCNGVR-----GEFDEIN 98
DB 136 GEDDSKETNDVVAQKEVEN-GSKEVTPCDSQRKDEANAGSEKVN 150
RESULT 11
C81232
50S ribosomal protein L24 NMB0153 [imported] - Neisseria meningitidis (strain MC58 sero-
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
C:Accession: C81232; D82004
A:Authors: H. Saunders, N.J.; Heidelberg, J.; Jettliffe, A.C.; Nelson, K.E.; Eisen, D.A.;
Tettelin, H.; Saunders, N.J.; Salzberg, S.L.; White, C.; Fleischmann, R.D.; Dougherty, B.A.;
Hickey, E.K.; Haft, D.H.; Vamathevan, J.; Gil, J.; Scarlato, V.; Maignan, V.; Pizzia, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.K.; Sarpola, R.; V
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81006; M01D:20175755; PM:10710307
A:Accession: C81232
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <NET>
A:Cross-references: GB:AE002373; GB:AE002098; NID:g7225359; P1CN:AAF50511.1; P1D:g722537
A:Experimental source: serogroup B, strain MC58
A:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.P.; Morel
Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mangill, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z3491.
A:Reference number: A81775; M01D:2022556; PMID:10761919
A:Accession: D82004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; P1CN:CAB93433.1; P1D:g737882
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: rplX; NMB0153; NMA0118
C:Superfamily: Escherichia coli ribosomal protein L24
Query Match 11.7%; Score 73; DB 2; Length 107;
Best Local Similarity 24.3%; Pred. No. 7.1;
Matches 25; Conservative 22; Mismatches 34; Indels 22; Gaps 4;

QY 3 NSLKKERVELDNCKSGNRGKPSSTFVR-TVTEFEVDFKILRRVHVVAIRTVAKVNGGV 62
DB 2 NKIKGRVVVVIAGKSKGKG-----QVAVLGDKVVEGVGVVVKR-HOKPNPMRJEGGI 56
QY 63 AEGFLPSKKKKRQNLGRSLDCNGVR--DGEFDEINRVGLQ 103
DB 57 STKEMP-----LDISNIALINPETNKADRVGIK 84
RESULT 12
C46622
hypothetical protein F18022.260 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48622
R:Boyan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24493
A:Accession: T48622
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-350 <HEV>
A:Cross-references: EMBL:DB
A:Experimental source: cultivar Columbia; BAC clone F18022
C:Genetics:
A:Map position: 5
A:Introns: 50/3; 113/1; 275/3
A:Note: F18022.280
Query Match 11.7%; Score 73; DB 2; Length 350;
Best Local Similarity 28.4%; Pred. No. 26;
Matches 27; Conservative 13; Mismatches 29; Indels 26; Gaps 4;
QY 7 KEERVELDNCKSGNRGKPSSTFVR-TVTEFEVDFKILRR 47
DB 220 KQEKTTIDESGSGVRGPGTPKTSTITQVRPVISVDEDFADDDSKMVLSEYAEGLN 275
QY 48 VHVATR-----TVAKVNGVVA--EGELPSKKKKRS 75
DB 250 ICEASSGSTSDKIAKVGNSVIEDNLMKKIEAS 314
RESULT 13
C47455
hypothetical protein T26112.40 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47655
R:Manfort, A.; Gasacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Jencke, K.; Mayer, K.F.
submitted to the Protein Sequence Database, February 2000
A:Reference number: Z24471
A:Accession: T47655
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2149 <MON>
A:Cross-references: EMBL:AL132954
A:Experimental source: cultivar Columbia; BAC clone T26112
C:Genetics:
A:Map position: 3
A:Introns: 882/1; 1043/3; 1133/2; 1320/2; 2069/1
A:Note: T26112.40
Query Match 11.7%; Score 73; DB 2; Length 2149;
Best Local Similarity 23.0%; Pred. No. 2e+02;
Matches 31; Conservative 27; Mismatches 55; Indels 22; Gaps 4;
QY 3 NSLKKERVELDNCKSGNR-----RGKPSSTFVR-TVTEFEVDE--FFK 43
DB 1330 SELKAATDLDTSQSDSNLANLVHPSLWPIILLSLKPS--TASESGDDLDPEVFMP 1387
QY 44 ILRRVHVVAIRTVAKVNGVVAEGELPSKKKKRSQNLGRSLDCNGVRDGEFDEINRVGLQ 103
DB 1388 FIMKCSOSNLRLVRVLASRALVGLVSNKLSQSLVLRITASTLPNSGAOGGSFNYLHGILLQ 1447

A1Status: preliminary
A1Molecule type: DNA
A1Residues: 1-464 <KUR>
A1Cross-references: GB:AF006641; NID:q13813145; P1DN:AAK40393.1; GSFD:GN00155
C1Genetics:
A1Gene: SS00076

```
Query Match:      11.6%   Score 72.5; DB 2; Length 464;
Rest Local Similarity 26.6%; Pred.No. 40;
Matches 33; Conservative 19; Mismatches 51; Indels 21; Gaps
Cy    5 LKREERVEEDNGSDGNRCKPSTEVYRTTEEEVDDEFFKILRRVHVATTVAKVNGVAF 64
DL          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cy    68 LDERSRVVAN-----NRAEPTTYYVKRQTEDDTPD-----IGDYIVTHNGITAN 114
DL          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cy    65 G-ELPSK---KKRKSONLGS:RNSLDCNGVRDSFEFDINRVGLQ---GLGLDUNCKPEPS 117
DL          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cy    115 DLELKXKYLKRKTIDSAIPLLL-DKTDWGNLEALKGILEQIKGSFALVGDKKNPR 173
DL          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cy    118 VSLS 12:
DL          : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: December 4, 2002, 09:26:07
Job time : 54 secs

```

QY 104 -GUGLNLCKPFFCS 117
      1 1 1 1
DB 1448 LGNLDTNGRLADN 1462

RESULT 14
SS0961
hypothetical protein: YNL166c - yeast (Saccharomyces cerevisiae)
NAlternate names: hypothetical protein N1706
CSpecies: Saccharomyces cerevisiae
CDate: 15-Feb-1996 #sequence revision
CRevision: 01-Mar-1996 #text change 19-Apr-2002
CAccession: S60961; S63118; S63908
R.Nasr, F.; Becam, A. M.; Herbert, C. J.
submitted to the EMBL Data Library, October 1995
A.Description: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 complete
tonic dystrophy kinase.
A.Reference number: S60958
A.Accession: S60961
A.Molecule type: DNA
A.Residues: 1-448 <NAS>
A.Cross-references: EMBL:X92517; NID:q1050783; PID:q1050787
R.Nasr, F.; Becam, A. M.; Herbert, C.
submitted to the Protein Sequence Database, April 1996
A.Reference number: S62967
A.Accession: S63118
A.Molecule type: DNA
A.Residues: 1-448 <NAS>
A.Cross-references: EMBL:Z71442; NID:q1302137; PID:e239539; PID:q1302138; MIPS:YNL166c
A.Experimental source: strain S286C
R.Nasr, F.; Becam, A. M.; Herbert, C. J.
yeast 12, 169-175, 1996
A.Title: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 complete
dystrophy kinase.
A.Reference number: S63805; MUID:96287653; PMID:8686380
A.Accession: S63808
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-448 <NAS>
A.Cross-references: EMBL:X92517; NID:q1302783; PID:CAA63273.1; PID:q1050787
A.Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
Genetics:
A.Gene: SGD:BN15
A.Cross-references: SGD:S000511C
A.Map position: 14L
Note: YNL166c

```

```

Query Match      11.6%; Score 72.5; DB 2; Length 448;
Best Local Similarity 27.5%; Pred. NO. 39;
Matches 30; Conservative 14; Mismatches 44; Indels 2; Gaps 5;

2Y 9RVEEDNGKSDGNGRKDPSTEVRTVTEEEVDEPKILRKRVHVATRIKAVKNGVAE-GEL 67
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 146 ENKEDTVN-DQNRGEADVVEIINNDQENEDKSDV-----EERRVKQGTENSEE 196
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

2Y 65 PSKKRRKRSQ--LCGRNSLDCNGVRGGEFEDEIRKVLQGGAGLQNTKPE 114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB 197 ESKSKSPMPQNTGGGNKKLDAELVLD-KFSSANK-----GLLTQFO 246
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESU:T 15
 H90140
 r:transcription regulator (xsbS) related protein [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C:date: 24-May-2001 #sequence:revision 24-May-2001 #text_change 24-May-2001
 C:Accession: H90140
 R:Sheng, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayez, M.J.; Chant-
 ong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi Ngoc, H.P.; Redder, R.
 arett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, C.
 submitted to GenBank, April 2001
 A:description: Sulfolobus solfataricus complete genome.
 A:Reference number: A90139
 A:Accession: H90140

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compend Ltd.

MM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 06:37:47 : Search time 53 seconds
(without alignments)
8275.405 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaaca.....ttacgaacttaagaaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

searched: 441362 seqs, 1533381 residues

total number of hits satisfying chosen parameters: 882734

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 9%
Maximum Match 100%
Listing first 45 summaries

Database : issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCRVUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103.2	6.1	7218	1	US-08-232-463-14
2	75	4.5	19124	2	US-08-487-826B-13
3	71.2	4.2	152331	3	US-09-128-555-16
4	71.2	4.2	176373	3	US-09-128-155-17
5	62.6	3.7	152331	3	US-09-128-555-16
6	58	3.4	6124	4	US-08-213-119B-3
7	56.8	3.3	289	4	US-09-057-005-17
8	56.8	3.3	289	4	US-09-244-795-17
9	55.6	3.3	291	1	US-07-942-723A-7
10	55.6	3.3	291	1	US-07-799-826C-7
11	55.6	3.3	291	1	US-08-074-275-7
12	55.6	3.3	291	1	US-08-480-366-7
13	55.6	3.3	291	2	US-07-952-277A-7
14	55.6	3.3	19124	2	US-08-487-826B-13
15	55.6	3.3	44453	4	US-09-146-053-5
16	54.2	3.2	5852	1	US-07-867-106-2
17	53.4	3.1	1956	4	US-08-555-906B-1
18	52.8	3.1	7379	4	US-09-341-587-5
19	52.4	3.1	1298	4	US-08-971-089-1
20	51.8	3.0	454	2	US-08-623-906A-6
21	51.6	3.0	72604	4	US-09-266-992-7
22	51.6	3.0	72604	4	US-09-657-474-7
23	50.4	3.0	1559	4	US-09-018-095A-7
24	50.2	3.0	2792	4	US-08-570-367C-1
25	50	2.9	1298	3	US-08-948-705-3
26	49.6	2.9	417	4	US-08-559-896B-3
27	49.6	2.9	1859	3	US-08-691-563C-46

Sequence 5, Appli
Sequence 4, Appli
Sequence 103, App
Sequence 103, App
Sequence 103, App
Sequence 10, Appl
Sequence 13, Appl
Sequence 595, App
Sequence 7, Appli
Sequence 7, Appli
Sequence 14, Appl
Sequence 1, Appli
Sequence 6, Appli
Sequence 13, Appli
Sequence 14, Appl
Sequence 15, Appl
Sequence 16, Appl
Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-06-232-463-14/C
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATOR: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/POCKET NUMBER: 30472/-14 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZ9pl-Fls
US-08-232-463-14

Query Match 6.1% Score 103.2 DB 1: Length 7218:

JS-09-128-155-16

Query Match: 4.2%; Score 71.2; DB 3; Length 152331;
Best Local Similarity 54.6%; Pred. No. 1.2e-06;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AATAAAGATGAGATAGAGAGATTCTGTGGAAGAAAGAGAGAGATGAGTGTG 262
DB 128751 AGACAAGAAAG 128694
QY 263 AACAAATTAAGAGATATCATGATATATTTATGAGAGTGTGTGAAGATTATTIAGAG 322
DB 128691 AGAAAG 128632
QY 323 AGGAG 382
DB 128631 AGAAAG 128572
QY 383 AAGATGAAG 442
DB 128571 AAG 128512
QY 443 ACAAAAG 462
DB 128511 AGGAG 128492

RESULT 4

JS-09-128-155-17/c
Sequence 17, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:

APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128.155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3)...(176373)
OTHER INFORMATION: n = A,T,C or G

JS-09-128-155-17

Query Match: 4.2%; Score 71.2; DB 3; Length 176373;
Best Local Similarity 54.6%; Pred. No. 1.3e-06;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AATAAAGATGAGATAGAGAGATTCTGTGGAAGAAAGAGAGAGATGAGTGTG 262
DB 144774 AGCAAGAAAG 144715
QY 263 AACAAATTAAGAGATATCATGATATATTTATGAGAGTGTGTGAAGATTATTIAGAG 322
DB 144714 AGAAAG 144655
QY 323 AGGAG 382
DB 144654 AGAAAG 144595
QY 383 AAGATGAAG 442
DB 144594 AAG 144535

QY 443 ACAAAAG 462
DB 144534 AGGAG 144515

RESULT 5

US-09-128-155-16
Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:

APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128.155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G

Query Match: 3.7%; Score 62.6; DB 3; Length 152331;
Best Local Similarity 51.2%; Pred. No. 0.0001;
Matches 146; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 173 ACATTGACGGAATTCAGCTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 232
DB 146914 AGATGGAG 146973
QY 233 ATGGAAG 292
DB 146974 AGAAG 147033
QY 293 TATGAGAGAGTGGTGAAGATTATTTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAG 352
DB 147034 AAGGAAG 147095
QY 353 ATGCTGAATCTGAAGAGAGATGAATTGTGTTAAAGATGAAGAGAGAGAGAGAGAG 412
DB 147094 AAGGAAG 147153
QY 413 CTAAGTCTCTGAAGAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
DB 147154 AAGGAAG 147198

RESULT 6

US-09-128-419B-3
Sequence 3, Application US/08213429B
Patent No. 633406
GENERAL INFORMATION:

APPLICANT: Inselburg, J. et al.
TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
FILE REFERENCE: J11-002CNP
CURRENT APPLICATION NUMBER: US/08/213,419B
CURRENT FILING DATE: 1994-03-14
PRIOR APPLICATION NUMBER: US 07/870,506
PRIOR FILING DATE: 1992-04-17
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0


```
US-08-074-275-7
Sequence 7, Application US/08074275
Patent No. 5468610
GENERAL INFORMATION:
  APPLICANT: Drs. Carl R. Merrill and
  APPLICANT: Michael H. Polymeropoulos
  TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
  TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
  NUMBER OF SEQUENCES: 9
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Lowe, Price, LeBlanc & Becker
  STREET: Suite 300, 99 Canal Center Plaza
  CITY: Alexandria
  STATE: Virginia
  COUNTRY: USA
  ZIP: 22314
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: DOS Text File
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/074.275
  FILING DATE:
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/07/707.501
  FILING DATE:
  ATTORNEY/AGENT INFORMATION:
  NAME: J.G. Mullins
  REGISTRATION NUMBER: 33073
  REFERENCE/DOCKET NUMBER: 717081
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 703 684 1111
  INFORMATION FOR SEQ ID NO: 7:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 291
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
  US-08-074-275-7

  Query Match 3.3%; Score 55.6; DB 1; Length 291;
  Best Local Similarity 52.6%; Pred. No. 0.00065;
  Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

  Cy 237 AAAAAGAGAGAGACATGTAGGTGAACAAAATAAGAGATATGATATATTTTATG 296
  Db 33 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92
  Cy 297 AGAGTGTGAAGATTTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 356
  Db 93 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 152
  Cy 357 TGAATCTGAAGAGATGAAATTTGTTAAAGATGAAGAGAGAGAGAGAGAGAGAG 416
  Db 153 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 212
  Cy 417 AGTCTCTGAAGAGATGAAAACAAACAAACAAACAAACAAACAAACAAACAAAC 466
  Db 213 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 262

RESULT 13
US-07-952-277A-7
Sequence 7, Application US/07952277A
Patent No. 5861504
GENERAL INFORMATION:
  APPLICANT: Drs. Michael H. Polymeropoulos
  APPLICANT: and Carl R. Merrill
  TITLE OF INVENTION: ELEVEN HIGHLY INFORMATIVE
  TITLE OF INVENTION: REPEAT POLYMORPHIC DNA MARKERS
  NUMBER OF SEQUENCES: 85
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Lowe, Price, LeBlanc & Becker
  STREET: Suite 300, 99 Canal Center Plaza
  CITY: Alexandria
  STATE: Virginia
  COUNTRY: USA
  ZIP: 22314
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible

US-08-480-366-7
Sequence 7, Application US/08480366
Patent No. 5721100
GENERAL INFORMATION:
  APPLICANT: Drs. Carl R. Merrill and
  APPLICANT: Michael H. Polymeropoulos
  TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
  TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
```


